Semi-Automated Annotation of Environmental Acoustic Recordings

Anthony Truskinger

BInfoTech (Hons) (Queensland University of Technology)

A thesis by publication in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

May 2015

§

Principal Supervisor: Prof. Paul Roe
Associate Supervisor: Dr Michael Towsey

Science and Engineering Faculty
Electrical Engineering and Computer Science
Queensland University of Technology
Brisbane, Queensland, Australia

This thesis is dedicated to my parents,
for never doubting that I could finish.

Keywords

- Acoustic Analysis
- Acoustic Sensing
- Analysis
- Annotation
- Audio
- Citizen Science
- Crowdsourcing
- Ecology
- Folksonomy
- Global Climate Change
- Linking
- Participants
- Participatory Analysis
- Recording
- Semi-Automated
- Sensors
- Spectrograms
- Tagging
- Taxonomy

Abstract

Title: Semi-Automated Annotation of Environmental Acoustic Recordings

Biodiversity monitoring is important for understanding the effects of climate and land use change. However, traditional biodiversity monitoring is a predominately manual process and hence the scale of monitoring is limited. Replacing manual fieldwork with acoustic sensors is an effective method to scale biodiversity monitoring over large spatiotemporal scales. After data is collected with sensors, the raw audio data must be analysed to produce interpretable results. Identifying the fauna that vocalise within the audio data is a common method of analysis. The data produced by fauna identification can be directly used to answer ecological questions.

Completely automated, high-accuracy methods for fauna identification in acoustic sensor data are promising but currently not feasible. Alternately, manual analysis is possible but inefficient. A compromise is a semi-automated approach: a methodology that combines the complimentary aspects of human analysts and computational resources. Human analysts have superior classification abilities, whereas automated computational resources are capable of working with data of massive scales. Analysts should be computationally supported for any data intensive task they undertake; this research investigated methods for supporting analysts who identify faunal vocalisations in the massive amounts of acoustic data collected by sensors.

This thesis is presented as a series of original research publications, modelled on the steps required to annotate faunal vocalisations in acoustic sensor data: detection, segmentation, and classification. Each of the publications is designed to make manual analysis more efficient for one of these annotation steps.

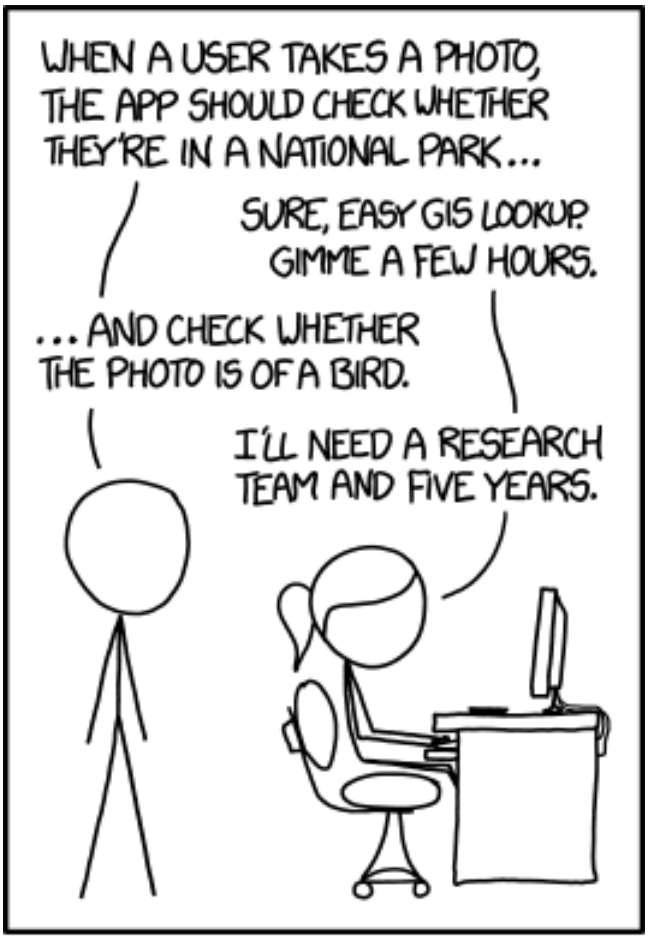
The first section of research (Chapter 4), rapidly scanning spectrograms, analysed the speed at which participants can detect acoustic events within static spectrogram images. It found that exposing a 24 second spectrogram image for as little as two seconds is enough time for analysts to decide if a koala bellow was present. This effectively reduced the time taken to do detection by a factor of 12.

The second section of research (Chapters 5 and 6) is a decision support tool for annotations. Typically, when classifying unknown acoustic events, analysts need to be able to recall, from memory, a large corpus of faunal vocalisations to be effective. The tool reduces recall requirements needed by analysts by suggesting possible species that may have emitted the vocalisation. To test the effectiveness of a decision support tool an experiment was setup using a dataset of 80 000 annotations with 400 types of vocalisations. The results of experimentation show that with basic

metadata features and a scale-tolerant algorithm, accurate suggestions can be presented for 48% of test cases.

The third section of research (Chapter 7), tag cleaning and linking, focussed on the last step of the annotation process: classification – specifically, applying a tag label (a class) to an acoustic event. This research aids analysts by repairing existing errors in a tag folksonomy. Repairing these errors allows the data generated by annotation to be used by ecologists, without first requiring laborious cleaning and normalisation. Additionally, the consistency gained from the automated cleaning, allowed the folksonomic tag data source to be linked to external taxonomic data sources. This linking allows richer data to be presented to analysts in future analysis tasks.

This thesis presents original research with the common theme of providing computer assistance to manual annotation methods in a faunal acoustic event annotation system. Assisting analysts increases their efficiency and allows more data to be analysed for a reduction in human resources. In combination, these publications make a significant contribution to the field of semi-automated faunal acoustic event annotation.



Randall Munroe
24/9/2014

http://xkcd.com/1425/

List of Publications

Publications that contribute directly to this thesis

This document is a thesis written by publication. Each of the following papers are presented as a chapter within this thesis.

1. **Truskinger, A.**, Cottman-Fields, M., Johnson, D., & Roe, P. (2013). *Rapid Scanning of Spectrograms for Efficient Identification of Bioacoustic Events in Big Data*. Paper presented at the 2013 IEEE 9th International Conference on eScience (eScience), Beijing, China. http://dx.doi.org/10.1109/eScience.2013.25
2. **Truskinger, A.**, Yang, H. F., Wimmer, J., Zhang, J., Williamson, I., & Roe, P. (2011). *Large Scale Participatory Acoustic Sensor Data Analysis: Tools and Reputation Models to Enhance Effectiveness*. Paper presented at the 2011 IEEE 7th International Conference on E-Science (e-Science), Stockholm. http://dx.doi.org/10.1109/eScience.2011.29
3. **Truskinger, A.**, Towsey, M., & Roe, P. (2015). Decision Support for the Efficient Annotation of Bioacoustic Events. *Ecological Informatics, 25*, 14-21. doi: 10.1016/j.ecoinf.2014.10.001
4. **Truskinger, A.**, Newmarch, I., Cottman-Fields, M., Wimmer, J., Towsey, M., Zhang, J., & Roe, P. (2013). *Reconciling Folksonomic Tagging with Taxa for Bioacoustic Annotations*. Paper presented at the 14th International Conference on Web Information System Engineering (WISE 2013), Nanjing, China. http://dx.doi.org/10.1007/978-3-642-41230-1_25

5. **Truskinger, A.**, Cottman-Fields, M., Eichinski, P., Towsey, M., & Roe, P. (2014). *Practical Analysis of Big Acoustic Sensor Data for Environmental Monitoring*. Paper presented at the 2014 IEEE Fourth International Conference on Big Data and Cloud Computing (BdCloud), Sydney, Australia. http://dx.doi.org/10.1109/BDCloud.2014.29

   Publication 5 is a primary author publication that is not part of main narrative of this thesis. The work was conducted during the author's PhD candidature and reflects the practical nature of developing bioacoustics software. It is closely related to the methodology chapter (Chapter 3) and is included in that chapter.

Publications indirectly associated with this thesis

The following are publications the author has contributed to as an author. These publications are ancillary research and their content is not detailed in this thesis.

1. Cottman-Fields, M., **Truskinger, A.**, Wimmer, J., & Roe, P. (2011). *The Adaptive Collection and Analysis of Distributed Multimedia Sensor Data.* Paper presented at the 2011 IEEE 7th International Conference on E-Science (e-Science).

   Contribution: Participated in the design and construction of the sensor network – particularly the software side. Relevance: This groundwork made the research in this thesis possible.

2. Duan, S., Towsey, M., Zhang, J., **Truskinger, A.**, Wimmer, J., & Roe, P. (2011). *Acoustic component detection for automatic species recognition in environmental monitoring.* Paper presented at the 2011 Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP).

   Contribution: Minor feedback on algorithm design and testing as a supporting researcher. Also contributed in writing the publication. Relevance: Important knowledge of automatic algorithms and the problems encountered when designing them. The author has experienced at firsthand the difficulties associated with automatic algorithm design.

3. Duan, S., Zhang, J., Roe, P., Wimmer, J., Dong, X., **Truskinger, A.**, & Towsey, M. (2013). *Timed Probabilistic Automaton: A Bridge between Raven and Song Scope for Automatic Species Recognition.* Paper presented at the Twenty-Fifth IAAI Conference.

   Contribution: Minor contribution in writing the publication. Relevance: A deep understanding of the software packages currently available for analysing acoustic data.

4. Zhang, J., Huang, K., Cottman-Fields, M., **Truskinger, A.**, Roe, P., Duan, S., . . . Wimmer, J. (2013, 3-5 Dec. 2013). *Managing and Analysing Big Audio Data for Environmental Monitoring.* Paper presented at the 2013 IEEE 16th International Conference on Computational Science and Engineering (CSE).

Table of Contents

List of Figures

List of Tables

List of Abbreviations

FFT Fast-Fourier Transform

IT Information Technology

ML Machine Learning

POI Point Of Interest

SNR Signal to Noise Ratio

UI User Interface

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

QUT Verified Signature

Signature:

Date: 21/05/15

Acknowledgements

To all the people who have supported me during the adventure that has been my PhD journey: I thank you for the consistent help, friendship, and encouragement you have provided me.

To my friends and family, thank you for putting up with the long hours, sporadic contact, and for being the guinea pigs for my experiments.

To my research group, thank you for the companionship. Doing a PhD with other research students is profoundly reassuring when the most frustrating parts of research get you down. In particular, special thanks go to Jason Wimmer and Mark Cottman-Fields. Both of these colleagues have consistently given me valuable feedback, support, encouragement, and advice. Without them, truly, this PhD would not have been possible.

To my supervisors, Paul Roe and Michael Towsey, my future as an academic will be possible because of you. Michael has consistently provided valuable feedback for some of the most complex parts of my thesis. He has endured hours of my attempting to explain my sometimes naïve methods and thought processes; the result is a thesis that demonstrates a level of rigor I would not have been capable of myself. I truly have learnt a great deal from him. Paul, my principal supervisor, has devoted hundreds of hours to managing the progress of my PhD, constantly encouraging me to question my assumptions and consider the real problems I was dealing with. Paul has encouraged me to work at my full potential, even during the times where I could barely work at all.

I would also like to thank the entities that supported me financially throughout my PhD. I am enormously grateful for the support I have received from the Australian Postgraduate Award and the Queensland University of Technology. I would also like to thank the Microsoft QUT eResearch Centre that was funded by the Queensland State Government under a Smart State Innovation Fund (National and International Research Alliances Program). Parts of this research were conducted with the support of the QUT Institute of Sustainable Resources and the QUT Samford Ecological Research Facility.

Professional Editor, Diane Kolomeitz, provided copyediting and proofreading services according to the guidelines laid out in the university-endorsed national policy guidelines. For more information, please refer to this link: http://iped-editors.org/About_editing/Editing_theses.aspx

Chapter 1
Introduction

Our environment is a precious resource, a complex series of ecosystems that exist as a set of careful balances between all forms of life on our planet. Research into flora, fauna, and their ecosystems is necessary to understand their interdependent relationships. Estimations of extinction rates are high (Thomas et al., 2004); to maintain ecological health and biodiversity, it is important to monitor the environment. Faunal monitoring is an important part of environmental monitoring.

Ecologists are the scientists tasked with studying the biological environment. Much of the work ecologists do is manual. Fieldwork requires human resources that are expensive and limited. Ecologists, like most scientists, try to scale their data collection methodologies. To scale data collection, ecologists are increasingly relying on technology to assist their research. As ecologists often are not technology experts, the result is inefficient use of technology.

The use of technology to support modern science problems is termed *eScience* (or cyber-infrastructure) and is particularly well suited for problems that involve big data or big compute tasks (Jankowski, 2007). By definition, eScience research is interdisciplinary. The research in this thesis takes form as eScience: by using acoustic sensors to assist ecologists, it is possible to massively scale ecological observations both spatially and temporally. This is done by detecting fauna that vocalise in audio recordings collected by sensors. Those vocalisations are provided to ecologists and they, in turn, use those vocalisations to make ecological inferences.

Using sensors as acoustic data recorders is just one possible method for scaling the monitoring of an ecosystem. Research has been conducted into monitoring systems for flora and fauna that use cameras and other sensors. Different methods of observation have different advantages.

Acoustic sensors are an increasingly common method for monitoring biodiversity. They can remain deployed in the environment for extended periods to record the sounds of the environment both passively and objectively. Sensors allow ecologists to scale data collection in both spatial and temporal dimensions. Sensor deployment, maintenance, or data collection are the only times humans resources are needed, resulting in cost-efficient data gathering.

Acoustic sensors can be used to monitor terrestrial or marine fauna but the equipment and approaches to monitoring each differs significantly. This research focusses on monitoring terrestrial fauna. To be detectable, fauna must emit some form of vocalisation. Therefore, any ecological inferences made rely on using vocalising fauna as a proxy for the health of their ecosystem. Analysis of the data is done to detect vocalisations, making it possible for ecologists to calculate metrics that estimate the overall biodiversity and health of the ecosystem from which the recordings were taken.

However, the actual identification of faunal vocalisations is difficult. Apart from the currently inadequate automated analysis methods, there are ranges of methodologies that integrate citizen scientists (interested participants from the wider community) as human analysts. These citizens analyse the recorded audio data by annotating acoustic events to identify vocalising fauna. Figure 1 shows an example of an interface that allows for an annotation style analysis of acoustic sensor data. The human-based analysis produces valuable data that has been used for several studies (Ellis, Fitzgibbon, Roe, Bercovitch, & Wilson, 2010; Wimmer, Towsey, Roe, & Williamson, 2013), yet, the analysis these participants conduct is time consuming and inefficient (Wimmer, Towsey, Roe, et al., 2013). This thesis addresses inefficiencies that were present in methodologies that use participants as analysts for detecting faunal acoustic events.

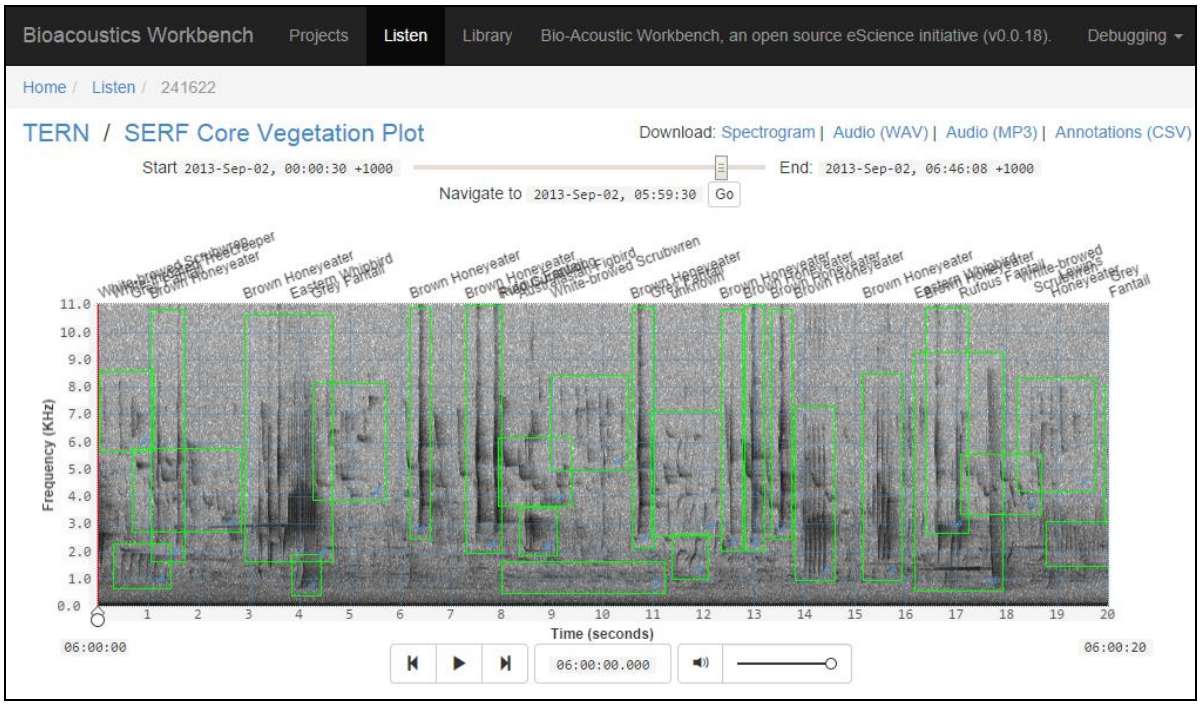


Figure 1 – An example annotation tool from the QUT Ecosounds software package.The tool can playback audio, visualise the audio, and allow a human analyst to create annotations (green rectangles) around interesting acoustic events. Annotations are labelled with tags.

Precisely, this research elucidates three methods for improving the efficiency of human analysts that annotate environmental acoustic events. As analysts exhibit diverse ranges of faunal identification skills, improvements in efficiency will enhance processing speed and relax requirements for human analysts to have expert training. The result makes the analysis task participable for arbitrarily skilled contributors.

1.1 Context

Acoustic sensors record data in real time – a sensor recording in the PCM WAVE format (16-bit, 2 channels, 22 050Hz) can generate up to 8GB of data per day. Thus, an abundance of sensors produce

massive amounts of audio data. Ecologists require the analysis of raw audio data but manual analysis is expensive. While purely automated methods for detecting species are not feasible, it is beneficial to analyse at least some of the acoustic data immediately, using other methods. Other methods produce data that is only a fraction of the volume, velocity, and complexity of that which a hypothetical, fully automated system could achieve.

This research has developed the concept of a semi-automated analysis methodology for bioacoustics that allows data analysis to begin immediately. The semi-automated methodology utilises analysts that are presented with audio and spectrograms (visualisations of audio). They analyse the audio for bioacoustic events using annotation tools to mark relevant sections of audio. The produced annotations are tagged with the species that emitted the vocalisation. The human analysts can be interested citizen scientists or ecologists.

Whilst the fully automated analysis of the environment is desirable, in general that problem is considered intractable. Therefor this research takes a different approach; it uses automation to reduce the effort required by analysts doing manual analysis.

Previous research has found human annotation speed to be slow (Wimmer, Towsey, Roe, et al., 2013). In this study, participants were instructed to annotate every unique faunal acoustic event they encountered. This required large amounts of time and effort from participants. Measurements show a time cost of a 2:1 ratio of analysis time to audio data. That is, it took analysts two hours to annotate every hour of audio data. The amount of effort needed to analyse significant amounts of audio is prohibitive. Even with a large workforce, the amount of work involved is off-putting for individual participants.

The result is an imbalance between the rate at which audio data is collected and the rate at which the data can be (nonautomatically) analysed. It was hypothesised that the efficiency of existing semi-automated methodologies and its participants could be improved.

1.2 Research Questions

The primary research question for this thesis follows:

How can automation improve the efficiency of manual analysis of faunal acoustic events in recorded acoustic data?

This question is addressed through three sub-questions:

1. Can the faunal event detection speed of analysts be enhanced?

2. Analysts must memorise large corpora of acoustic events to be effective; can this requirement be relaxed or reduced?
3. Can human generated folksonomies used to tag acoustic events be mapped back to taxonomies?

1.3 Significance of Study

This research contributes a set of methods for enhancing the efficiency of human analysts that are annotating bioacoustic events in a semi-automated analysis system. These improvements target different parts of the annotation process and address the three research sub-questions of this thesis. They are:

1. a simple method for identifying acoustic events via rapid-scanning spectrograms
2. an annotation suggestion algorithm that outputs a ranked list of potential matching vocalisations
3. heuristics for ensuring tags in the annotation folksonomy are consistent, correct, and can be mapped to species taxonomies

Improving analyst efficiency addresses the collection/analysis imbalance present in acoustic sensing. Without analysis of the data, ecologists are limited in what they can infer. Given the rate of data collection, unless more audio can be analysed, there will be no significant output of data for ecologists.

Analysis of audio with human participants, means annotation data can be produced sooner than if a completely automatic analysis was relied on. Eventually, high-accuracy, completely automated faunal event analysis will, for any species (in any ecological region, with any amount of background noise), be viable – with little or no human input. Currently though, a completely automated solution is considered intractable (the literature review covers this topic in depth).

1.4 Limitations of Study

Importantly, this research does not explicitly consider analysts' abilities. The research conducted includes data generated by both expert and novice analysts. The research goal is a general solution for all participants, despite their skill; generalised solutions were investigated rather than skill-specific solutions (which would have only targeted experts or novices).

The accuracy of annotations is a metric used to evaluate the correctness of the data output by this research. An annotation is considered correct if a participant correctly annotated an acoustic event. Guaranteeing data accuracy is also known as verification. Verification is an important part of analysis systems that use human computation and has been proven effective in its own right.

This research concentrates on improving the supporting technologies for semi-automated annotation. These processes occur before verification is required. Improving the efficiency of annotation consequently increases the probability of a correct identification. Part of this research includes ensuring there are no errors in the folksonomy of tags used for annotations. This is not the same as verification; just because the tags applied to an annotation are valid species names, it does not ensure they are correctly applied.

Additionally, this research does not plan to implement all methods of semi-automation into one experiment. Prototyping individual, isolated, components, is necessary to limit scope and complexity. Integrating all methods researched in this thesis would a) not constitute a well-designed experiment outright and b) require a large amount of programming work. As the research question will be tested on production-grade software, it is not planned to make any significant changes to the code base that would require complete and professional implementations – that work is considered outside the scope of a research problem.

While not purposely limited to one ecological scope, the outcomes of this research should be assumed limited to the ecological scope of the data collected. Preferably, this research would be applicable to ecological environments globally; however, this is unfeasible for two reasons. This research relies on data collected primarily from one source, resulting in a limited geographical subsample. Secondly, faunal vocalisations demonstrate variation throughout different geographical regions (Catchpole & Slater, 2008) – one technique that works in one region may fail in another. With these considerations in mind, this research is limited to ecological areas where the data was readily available. Examples of these areas (all located in Australia) include: the general Brisbane area, the QUT Samford Ecological Research Facility, Groote Eylandt in North Queensland, as well as St Bees Island located off Queensland coast near Mackay. Lastly, although none of the research in this thesis is specifically limited to Aves, over 90% of the acoustic events analysed in the available data were produced by Aves. There are examples of non-Avian species in the available data (including koalas, insects, mammals, and frogs); however, the research in this thesis has effectively only been tested on Avian vocalisations.

1.5 Thesis Structure

This thesis is presented by publication of papers according to the appropriate regulations for QUT PhD students. These regulations state can be found in Appendix A – QUT Thesis by Publication Regulations.

This thesis includes four core publications and one ancillary publication; all are peer reviewed and published. Each publication stands on its own as individual work. When combined, the core

publications form a cohesive narrative that addresses the research questions of this thesis. These publications each map to a chapter of this thesis. The core chapters of this thesis develop the concept of computer-assisted annotation of faunal acoustic events. There are three steps required to create an annotation: detection, segmentation, and classification.

Before the papers are presented, a literature review (Chapter 2) discusses, reviews, and compares literature relevant to the thesis. The *Background and* Methodology chapter (Chapter 3) describes the background, methodology, and development of the software artefacts associated with this research. This methodology chapter also includes an ancillary publication that further elucidates the general methodology employed for environmental acoustic monitoring.

The first major chapter details a method for increasing the speed of acoustic event detection by rapidly showing a series of static spectrograms to a participant. The experiment measures participant accuracy against varying exposure speeds in order to determine how fast visual acoustic event detection can occur. This chapter (*Rapid Scanning* of *Spectrograms,* Chapter 4) addresses sub-research question 1.

The second major chapter details a prototype of an analyst-oriented implementation of *FELT* (Find Events Like This). The aim of the FELT tool is to suggest to a participant what classification might be appropriate for an acoustic event that has been identified in a spectrogram. This chapter (*A Prototype Annotation Suggestion Tool*, Chapter 5) addresses sub-research question 2.

The third major chapter extends the FELT idea to implement the suggestion tool in full. This chapter presents improvements in accuracy achieved while scaling the input training data for the tool. This chapter (*Decision Support for the Efficient Annotation of Bioacoustic Events*, Chapter 6) also addresses sub-research question 2.

The last major chapter addresses data quality problems within tag data for already created annotations. The tag data generated by participants demonstrated a variety of errors that needed to be fixed before ecologists could make use of the data. This chapter (*Tag Cleaning and Linking*, Chapter 7) addresses sub-research question 3.

Finally, the thesis conclusions are presented in Chapter 8.

Figure 2 – Thesis overview

1.6 Ethics

All activity undertaken as part of the research conducted for this thesis occurred under the following ethics policies.

The ethics policy of the author's research group applied to any research that was general and did not involve participants. Examples of this kind of research include data analysis, creating or designing algorithms, running automated analyses, designing interfaces, and any other programming.

Any research that involved participants external to the research group was covered under an explicit ethics agreement. Ethics approval was sought from the QUT Ethics Committee as a *low-risk ethics application*. The ethics application was approved with the approval number of **1200000307** on the 18th June 2012 and was valid through to the 18th June 2015.

A copy of the ethics application approval email and cover sheet are included in the appendices.

Chapter 2
Literature Review

This literature review presents a comprehensive analysis of existing research related to this thesis. The review begins by introducing general concepts to explain or support subject matter at the base of the research topic. The review then discusses bioacoustics by presenting literature on motivations, related projects, collection methodologies, and existing analysis techniques. This discussion of bioacoustics is followed by examples of semi-automated analysis. Lastly, the concepts of tagging and existing research are discussed.

This chapter provides a broad summary of related work. In addition, since each paper is standalone work, they will also discuss related work. This chapter is designed to provide an overall summary of work related in the areas of human classification skill, bioacoustics, and tagging.

2.1 General concepts

This thesis is research positioned within several overlapping topics; this section of the literature review will briefly define some of the interconnected concepts that affect this thesis.

2.1.1 E-Science

eScience (electronic science, enhanced science, cyberscience, or cyber-infrastructure) is defined as using technology to support modern science problems, particularly those that involve big data or compute problems (Jankowski, 2007). eScience is by definition interdisciplinary.

The term e-Science was coined by John Taylor in 1999. eScience has roots in European research institutions that focus on the natural sciences. The original definition for eScience was restrictive: It was defined over just a few areas of computationally intensive IT research that intersected with other sciences. Big Data, distributed computing, and grid computing were the focus of eScience research groups. However, the definition has broadened to now include many modern technologies associated with big data scale scientific methodologies.

2.1.2 Citizen Science

Citizen scientists are *"volunteers that participate as field assistants for scientific studies"* (Cohn, 2008, p. 2). Citizen science involves everyday citizens with professional scientific projects. Often the volunteers involved are not currently or have never been professional scientists but rather are enthusiastic amateurs. Citizen science is a type of crowdsourcing methodology.

Citizen Science has shown promise for research projects (R. Sullivan, 2009), in that citizens can devote often-precious resources like time and effort to them. When reviewing one of the projects in the case study, Sullivan states that the project's scientists were impressed by the dedication of the citizens working for it: "*Volunteer groups are very keen to produce robust, rigorous, properly*

collected information that feeds into something bigger, and has a significant impact…"(R. Sullivan, 2009, p. 12).

The Galaxy Zoo project is an example of a successful citizen science project (Galaxy Zoo, 2010). This project employs a Crowdsourcing model (using masses of ordinary citizens to work on a problem) in order to process large amounts of data. Galaxy Zoo uses its community to classify the morphology of images of different galaxies. This project is a great example of citizen science because is it utilises everyday citizens (from amateur astronomers to children), who are interested in astronomy, to make a marked contribution to the scientific field. Importantly, Galaxy Zoo's contributors do not collect data; they only validate and classify it.

In other citizen science projects, participants contribute both by analysing data (Galaxy Zoo: http://www.galaxyzoo.org) and collecting and contributing data (eBird: http://www.ebird.org). Given the varied background of citizen science participants (ranging from amateur enthusiasts to experienced scientists), there are significant challenges to be overcome with citizen science projects (Cooper et al., 2009). One of the foremost challenges is establishing the skill level or *reputation* of the participant performing the collection or analysis task. To achieve this, many citizen science projects utilise reputation management to classify participants and to establish the credibility of their contributions.

Galaxy Zoo is a classic example of this approach, with over 250 000 active users helping to classify galaxy types according to their shapes (Galaxy Zoo, 2010). The identification of galaxies is done automatically but the complex classification task is deferred to humans. Galaxy Zoo provides users with initial training and then tests their abilities. Verification through repeated classification of the same galaxy by multiple users ensures consistency and accuracy (Lintott et al., 2008). The data of citizen science projects is contributed by volunteers. Due to most having little or even no scientific training, the quality of contributed data is not guaranteed. Galaxy Zoo and other citizen science projects apply the concept of reputation management to their contributors, to weight the value of each user's contribution (Abdulmonem & Hunter, 2010; Burke et al., 2006; Huang, Kanhere, & Hu, 2010; Reddy et al., 2008).

2.1.3 Spectrograms

A *spectrogram*, or sonogram, is a visual representation of the spectrum of frequencies for sound data (Haykin, 1991). A spectrogram can visualise any stream data, not just sound data, as a time/frequency graph. In the context of acoustics, spectrograms allow for the recognition and association of visual patterns with acoustic signals. These graphs usually show a progression of time

along the x-axis and the frequency along the y-axis. Intensity is represented by colour, shade, or in the case of 3D representations, as height (projected along the z-axis).

Spectrograms are essentially the application of a *fast Fourier transform* (FFT) to a shifting window of samples from a larger data stream. The results of the transform are a set of frequency bins and their relative intensities for the time from which the subset of samples was taken.

Several examples of spectrograms are shown in Figure 3 along with a comparative waveform visualisation. Just the waveform visualisation shows only one dimension of data. The information shown in a waveform is limited compared to a spectrogram. The audio data was sourced from a file recorded on the 7th of October 2011, beginning at 06:00:00, taken from the Cornubia Wetlands Project, a project that collected data from the Cornubia Wetlands located in Logan, Queensland, Australia. The data shown is a 12s extract.

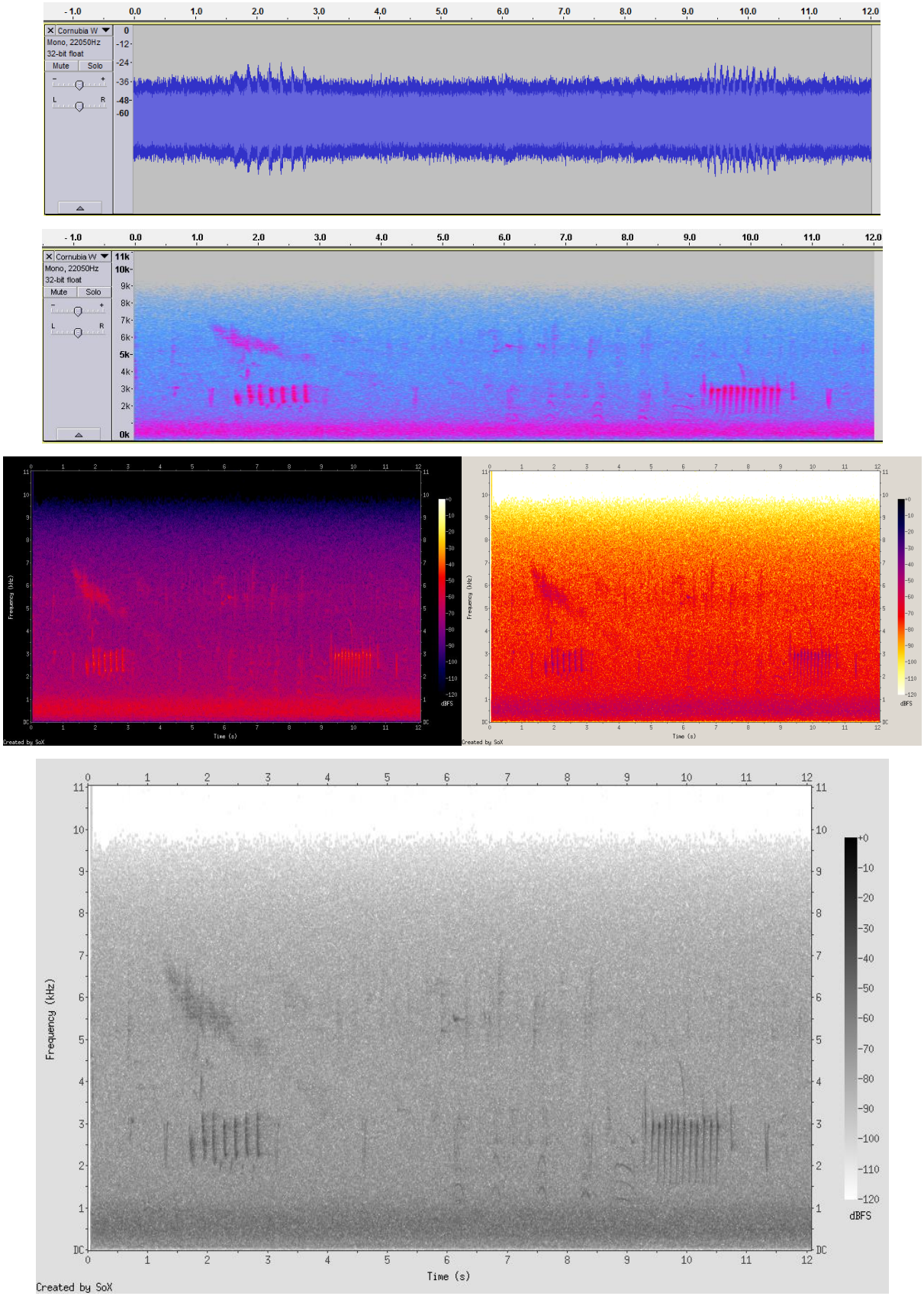


Figure 3 – Several examples of audio visualisation generation. In order from top to bottom: 1) a waveform produced by Audacity (an audio recording and manipulation tool); 2) the equivalent spectrogram (colour) generated by Audacity; 3) & 4) colour spectrograms produced by SoX (an audio manipulation tool); 5) the monochrome SoX spectrogram commonly used by the QUT Ecoacoustics research group.
References: (Audacity Team, 2013; Bagwell, 2013)

2.1.4 Machine Leaning

Machine Learning uses artificial intelligence research to allow algorithms to adapt to new situations or to more intelligently respond to stimulus (Michalski, Carbonell, & Mitchell, 1985; Mitchell, 1999). Machine Learning (ML) has particularly useful applications for classifying instances in large sets of data. This section introduces terms used elsewhere in this thesis.

In Machine Learning, the term '*classification*' refers to a procedure for assigning a given piece of input data into one or more categories. Sokal (1974) defines classification as: "*the ordering or arrangement of objects into groups or sets on the basis of their relationships. These relationships can be based on observable or inferred properties.*" Sokal breaks classification into two categories: monothetic and polythetic. Monothetic classifications are possible when there is one attribute present in all instances that distinguishes them. Polythetic classifications occur when more than one attribute is used to describe the differences between instances. Polythetic models are most common.

An algorithm that implements classification is called a classifier. The number of available classes can be predefined (typically for supervised ML algorithms) or created dynamically (typically for unsupervised algorithms).

An input datum is termed an instance. Categories are termed classes (or labels). An instance is described by a vector of features, which together constitute a description of all relevant discriminatory characteristics (attributes) of the instance. The range of valid values and features is called an input space. Each feature is equivalent to a dimension; for example, three features represent a three-dimensional problem. Three-dimensional and sometimes four-dimensional input spaces can be visualised; however, high dimensionality problems are also common. These problems cannot have their entire input space easily visualised, which makes understanding the data difficult for a human analyst.

Test data is a set of instances that are to be evaluated in an experiment and assigned a class. In the experiment, typically in order to evaluate correctness, the output features (labels or classes) are pre-filled. This is called labelling, encoding, or seeding the input features. The pre-filled data is ignored by the experiment until it is time to test the correctness of the results (the classifications). The labelled test data is known as the 'Gold standard'. Training data is the term used on the set of instances provided to the learning part of a ML algorithm. Training data is typically only used for supervised solutions. It is important that training and test instances are disjoint sets.

The *ground truth* should not be confused with a *gold standard*; a gold standard is the best available test (Versi, 1992) whereas the ground truth is a technique used to determine the right answer (e.g. verification) (McClatchie, Thorne, Grimes, & Hanchet, 2000).

Classification is a term used for supervised ML problems. Supervised procedures learn to classify new instances based on input training data. Classification can be binary or multiclass. Binary is the simplest and most common – the point of the algorithm is to determine if the instance belongs to a class or not. The results of binary classifiers can be represented by a confusion matrix (Table 1).

Table 1 – The confusion matrix of a binary classifier

		Condition (as determined by "Gold standard")		
		Condition Positive **P**	Condition Negative **N**	
Test Outcome	Test Outcome Positive	**True Positive TP**	**False Positive FP** (Type I error)	Positive predictive value = $\frac{\sum TP}{\sum \text{'Test Outcome Positive'}}$
	Test Outcome Negative	**False Negative FN** (Type II error)	**True Negative TN**	Negative predictive value = $\frac{\sum TN}{\sum \text{'Test Outcome Negatives'}}$
		Sensitivity = $\frac{\sum TP}{\sum P}$	Specificity = $\frac{\sum TN}{\sum N}$	

For supervised problems, over fitting occurs when the training model fits the training data too well. As a result, the performance of the classifier when applied to test data may suffer. Ideal training data should represent a well-distributed subset of a population. Commonly, that subset may not actually be representative of the general traits of the population. In that case, the assumptions made during training will affect classifier performance when used on a full population of data.

Unsupervised procedures do not rely on labelled training data. Clustering is an example of an unsupervised method that collects data into classes based on some measure of inherent similarity. As this thesis does not make use of any unsupervised methods, they are not elaborated on further.

Classification and clustering are examples of the more general problem of pattern recognition, which is the assignment of some sort of output value (a label) to a given set of input values. A common

subclass of classification is probabilistic classification. Algorithms of this nature use statistical inference to find the best class for a given instance. Unlike other algorithms, which simply output a "best" class, probabilistic algorithms output a probability of the instance being a member of each of the possible classes. The highest probability is usually used to select the best class. Because of the probabilities output, probabilistic classifiers can be more effectively incorporated as components into larger machine-learning tasks.

The following diagram (Figure 4) shows the basic workflow for a supervised ML algorithm. The training data is used by the training algorithm to form a classification model. The classification model is a representation of the training data that is a more compatible format for the classifier. Some methods do not make used of a training algorithm, in which case the classification model is synonymous with the training data. Classification takes each instance from the test data, and compares it to the classification model with the classifier algorithm. The result of classification is a set of labelled test instances.

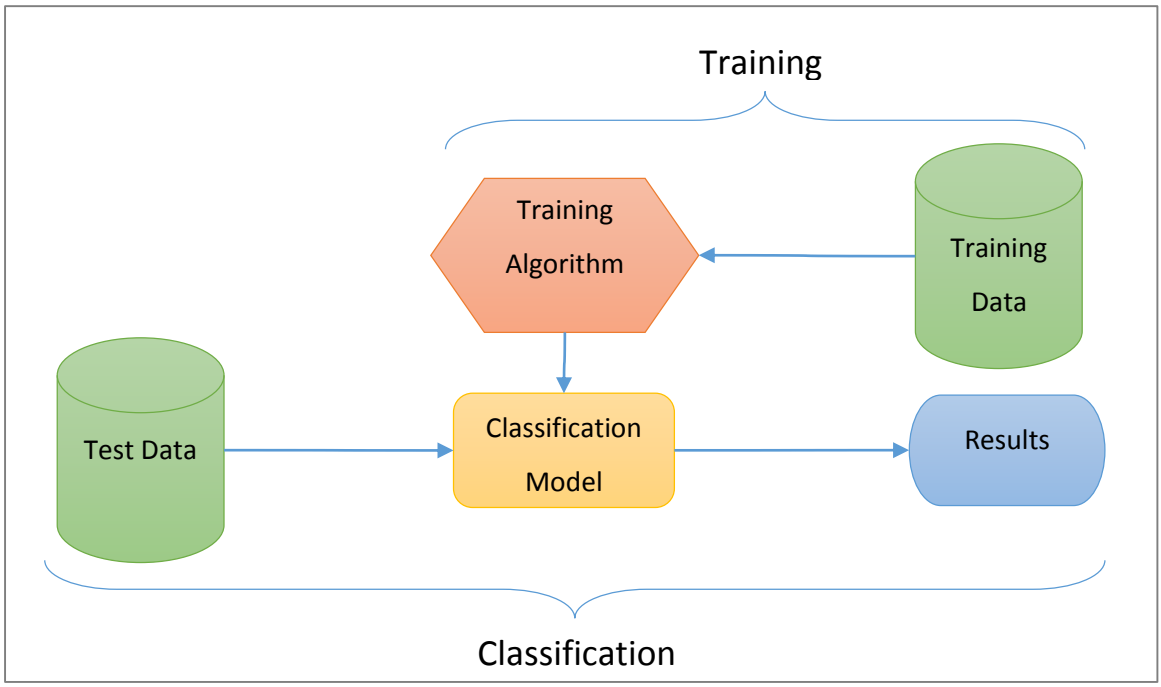


Figure 4 – A generalisation of supervised machine learning algorithms

2.1.4.1 Similarity Search

Similarity searches are designed to take advantage of the large amounts of data collected by databases to retrieve subsets of data that are similar to an input query (Gionis, Indyk, & Motwani, 1999; Zezula, Amato, Dohnal, & Batko, 2006). Similarity searches are designed not to classify instances but rather, the distances between them. The term *similarity search* covers a broad range of problems including facial recognition, audio matching, and more generalised database instance

searching (Chávez, Navarro, Baeza-Yates, & Marroquín, 2001). Many similarity search methods ensure that all input features are transformed into a metric space. Pattern retrieval is a term sometimes used for a sub class of similarity searches associated with image searching/retrieval (Ma & Manjunath, 1994).

Shazaam is a commercial music matching service that utilises similarity searches (Wang, 2006). Users of the service record short fragments of a song that they want identified on their mobile devices. The sample is sent to Shazaam as a query and then matched against a massive database of known patterns. Shazaam extracts features from source and query audio by detecting points of interest in the audio. The combination of the distributions of these points of interest, form a unique hash that can be easily indexed. Using hashes to encode feature sets has been applied successfully by others (Gionis et al., 1999).

2.1.4.2 Recommender Systems

Recommendation systems extend the concept of similarity searches. Recommender systems are a type of information filtering system that rates or predicts content a user wants (Ricci, Rokach, & Shapira, 2011). Most common examples of recommender systems are applied to content like large bodies of text, metadata for movies/music, and movie/music data itself. They are designed to assist users by delivering targeted content to those users who lack the personal experience needed to evaluate the massive number of choices they can make. Recommender systems have been used successfully in large-scale examples. Amazon's book selection, Pandora's radio feature, and Netflix's media suggestions, are all good examples of content-based recommender systems. Recommender systems have an advantage over searching, because the search process is automated and abstract hard-to-search-for concepts become suggestible automatically.

Recommender systems use profiles of their users that measure features such as what they like, dislike, and other demographics (Schein, Popescul, Ungar, & Pennock, 2002) – the queries to recommender do not radically change on every request. The concepts of recommender systems are similar to those of similarity searches. However, recommender systems search for user preferences – they are intentionally biased to return the best result for the user. As such, the features and techniques used are optimised in ways that are not generalisable.

2.1.4.3 MFCCs

MFCCs are a common feature extraction method used in human speech recognition (and other ML techniques). A Mel-filter is used as part of MFCCs for syllable detection (Bridle & Brown, 1974). MFCCs are extracted by applying a Mel (logarithmic) scale to the frequency domain of an FFT, then by taking the logs of the powers for the Mel frequencies, and then applying a discrete cosine

transform. The syllables that form human speech often take form as static harmonics that are well described by MFCCs. MFCCs are susceptible to noise (Tyagi & Wellekens, 2005). After the MFCCs are produced (as feature extraction), they are then passed to a classifier. Hidden Markov Models are common classifier used methods (Scharenborg, 2007).

The use of MFCCs in faunal acoustic event classification is discussed later (section 2.4.2).

2.1.5 Classification Abilities of Humans

The ML field is improving but humans are still generally better at classification tasks than machines. The literature in this section will demonstrate that when comparing machines to humans:

- Humans are better classifiers and recognisers of visual data
- Humans are better classifiers and recognisers of auditory data
- Humans can apply context and intuition to the classification process
- Humans are adept at pattern matching
- Humans can adapt to noisy, corrupted, or poor quality data

However:

- Machines are better at data mining (discovering patterns in masses of data)
- Machines are superior classifiers for purely numerical data
- Humans are inherently biased
- Humans can make mistakes, be inconsistent, get tired, or become bored
- Importantly, humans find it difficult to identify what they see when the number of classes is large. They can tell if two instances are similar but cannot necessarily name (classify) each instance.

The general goal of ML is to enable computers to adapt and learn from their environments just as humans do (Alpaydin, 2004). To learn and adapt, humans have evolved a large brain with a highly developed visual cortex – sight is the most used of the senses (Allman, 2000). Thus, it is ideal to rely on human vision to distinguish objects.

ML can solve problems that scientists have traditionally considered intractable for machines (Sokal, 1974). The minimum length set of algorithms (e.g. travelling salesman) are among the examples given. Another advantage of machine learning is the ability for an algorithm to classify many entities at once, using many more features than a human can. Advantages to classifying data include ease of manipulation, easy retrieval of information, and description of structure and relation to other objects so that general statements can be made (Sokal, 1974).

The contrast between what humans and machines are capable of is highlighted by Shneiderman (2003). They present useful information on integrating automated tasks into a user interface as seen in Table 2.

Table 2 – A comparison of the abilities of humans and machines (Shneiderman, 2003, p. 79)

Humans Generally Better	**Machines Generally Better**
Sense low-level stimuli	Sense stimuli outside human's range
Detect stimuli in noisy background	Count or measure physical quantities
Recognize constant patterns in varying situations	Store quantities of coded information accurately
Sense unusual and unexpected events	Monitor pre-specified events, especially infrequent ones
Remember principles and strategies	Make rapid and consistent responses to input signals
Retrieve pertinent details without a priori connection	Recall quantities of detailed information accurately
Draw on experience and adapt decisions to situation	Process quantitative data in pre-specified ways
Select alternatives if original approach fails	Reason deductively: infer from a general principle
Reason inductively: generalize from observations	Perform repetitive pre-programmed actions reliably
Act in unanticipated emergencies and novel situations	Exert great, highly controlled physical force
Apply principles to solve varied problems	Perform several activities simultaneously
Make subjective evaluations Develop new solutions	Maintain operations under heavy information load
Concentrate on important tasks when overload occurs	Maintain performance over extended periods of time
Adapt physical response to changes in situation	

2.1.5.1 Specific Studies

Sternberg (1966) published a paper that establishes the rate of symbol processing in the average human. His results suggest that humans can process a stream of symbols in order to pick out one, at a rate of 25-35 symbols per second (static image, low exposure). This paper supports the scanning theory, which details how a user might visually scan through data.

Feyyad (1996) discusses the topic of knowledge discovery in databases. Feyyad focusses on the "torrent of available data" and methods for making use of the data. They advocate removing the human-in-the-loop components out of analyses because there are generally too much data (in too many dimensions) for a human to adequately process. However, in advocating the removal of humans, the author also details what advantages are lost without human analysts. Feyadd credits humans with quick and correct decision making, especially when the data is visualised. Feyadd describes humans as good classifiers because they can identify useful low-level features in order to make an intuitive decision. Additionally, with visual representations, Feyadd credits the highly evolved human visual system for its ability to spot interesting details. However, they state that humans are not good at the data-mining task – i.e. finding useful information in masses of data.

A project dedicated to detecting lymphocytes in human tissue tested what was more accurate: human experts or automatic recognisers (Nattkemper, Twellmann, Ritter, & Schubert, 2003). The humans analysed images (micrographs) and the automatic algorithm used a neural network to classify the data. For images of good quality, the algorithm was equivalent in accuracy to a medium-skilled expert. In images with noise, some human experts outperformed algorithms. However, it was found that the algorithm is quicker to classify than the experts are. The paper notes that using an artificial neural network as its classifier (quick to use, slow to train) is quicker than using human analysts.

Another paper conducts an experiment comparing machine to human performance for classifying samples of Dinoflagellates (*Dinoflagellata*) (Culverhouse, Williams, Reguera, Herry, & González-Gil, 2003). Interestingly, in this case humans often do not perform adequately, leading to investigation into automated methods of classification. This paper acknowledges the existence of experts and associates them with a higher skill level than ordinary analysts. It then explains why the standard rules used for encoding taxonomy are not useful in an automated system. This is because taxonomies are defined by features that are naturally described well by humans and not by machines. The paper also defines a concept of consistency; to be viable as a classifier, a human must be accurate and consistent. Unsurprisingly, the authors found experts to be more consistent than amateurs. The paper also touches on the many forms of bias that humans encounter when trying to classify. Despite all the negative attributes this paper associates with human classifiers, it states that the automatic methods they tested performed no better. The result: a hybrid system was recommended.

Human Interactive Proofs (HIPs) are small reverse Turing tests. In this case, an HIP that uses handwriting for the proofing mechanism is a good example for detailing the superior classification abilities of human participants (Rusu & Govindaraju, 2004). An HIP is a challenge mechanism designed to test if the interactant is human. This paper adapts the CAPTCHA technology to test a new idea. CAPTCHAs verify the human interacting with the system, by encoding a sequence of characters into an image that is then distorted via various techniques. The result is an image that cannot be decoded by state-of-the-art artificial methods but is still relatively easy to decode for a human. CAPTCHAs rely on the superior classification ability of humans when compared to machines. Using handwriting as opposed to printed text as the source data was tested to strengthen the CAPTCHA concept (Rusu & Govindaraju, 2004). The results found that for machines, the best OCR has the equivalent performance of a 1st year primary school student. The paper states that character recognition has improved, but falters easily when noise is introduced. For additional security, the

authors artificially introduce noise that further decreases the chance of an algorithm performing well (whilst maintaining human readability). The types of noise they introduce are detailed:

> 1. *Noise: Add lines, grids, arcs, circles, and other "background noise"; Use random convolution masks, and special filters (i.e. multiplicative/impulsive noise, blur, spread, wave, median filter, etc)*
> 2. *Segmentation: Delete ligatures or use letters and digits touching with some overlap to make segmentation difficult; Use stroke thickening to merge characters*
> 3. *Lexicon: Use lexicons with "similar" entries, large lexicons, or no lexicons; Use words with confusing and complex characters such as "w" and "m"*
> 4. *Normalization: Create images with variable stroke width, slope, and rotations; Randomly stretch or compress portions of word image*
>
> *(Rusu & Govindaraju, 2004)*

The artificially introduced noise types are all similar to noise that could be seen in an ordinary spectrogram of an acoustic signal. For instance: random noise frequently occurs from a variety of sources; environmental conditions or low SNR can easily result in the loss of fine detail (like removing ligatures) and signals can often be segmented; similar looking /sounding vocalisations are frequent; and variation of vocalisation shape, duration, and bandwidth is known to vary over large distances. All of these sources of noise are added to the HIP signal to prevent machines from classifying components resulting in a set of images that have similar types of noise to acoustic sensor data. Due to the similarities in concepts, it can be seen why humans are often better classifiers of vocalisations, especially when provided with a spectrogram.

The performance of consonant recognition was measured by Sroka and Braida (2005). In addition to testing participants, the study used various machine learning algorithms to recognise consonants in a stream of audio. Importantly, as well as clear audio, the authors simulated noisy signal by adding in various types of noise. For signals that were subject to low or high pass filters, the algorithms worked as well as humans. When speech-like noise was added to the signal, the humans outperformed the algorithms. The paper also listed the then current state of speech recognition techniques:

> *Despite significant advances in Automated Speech Recognition (ASR) systems, performance at human levels has not yet been attained. Human recognition results provide proof that continuous speech can be recognized more accurately than the best current ASR systems.*

> *(Sroka & Braida, 2005)*

Scharenborg (2007) reviewed the gaps between two related areas of speech recognition (human speech recognition and automatic speech recognition). In this discussion, they list some characteristic differences between human and machine classification ability. Several other sources (Scharenborg, 2007, p. 5) are cited and it is concluded that the performance difference between human and machine classification increases when signal-to-noise ratio is low. It is also reasoned that pure performance should not be the only consideration; humans are better at adapting to non-stationary noise in the signal – a problem many automatic approaches do not account for. The authors reinforce the human ability to apply context and wisdom (prior knowledge) to the classification task (knowledge of the environment, world, topic, etc...). Even when human intuition cannot be applied, for example by removing all contextual information and asking the human to recognise nonsense words, humans still perform better than machines. In summary, the authors estimate the superiority of humans as classifiers as an order of magnitude over that of a machine. Additionally, humans are particularly capable of adapting to noise, signal rate, signal style, and overlapping signals.

2.2 Bioacoustics for Environment Monitoring

In the ecology research field there is much interest in recognising faunal events from audio data. *Bioacoustics* is the study of the emission, propagation, and reception of sound by fauna (Bradbury & Vehrencamp, 1998). Through bioacoustics the monitoring of species, their relationships, and biodiversity is possible; monitoring is important for providing insight into an ecosystem (Hu et al., 2009; Mason et al., 2008; Porter et al., 2005). Fauna is identified through vocalisations recorded using hardware or technology-assisted methods (S. Brandes, 2008; Mason et al., 2008; Planitz, Roe, Sumitomo, Towsey, Williamson, & Wimmer, 2009).

Much research has been done that links ecological condition to acoustic sound scapes. For an ecological example, the work by Tucker, Gage, Williamson, and Fuller (2014) links ecological condition (evaluated manually) to the relative soundscape power to reflect ecological conditions like bird species richness. Their conclusion is that acoustic monitoring is cost effective.

Using sensors can help remove some of the more common methods of bias introduced by human sampling techniques (S. Brandes, 2008; Mainwaring, Culler, Polastre, Szewczyk, & Anderson, 2002; Wimmer, Towsey, Planitz, Williamson, & Roe, 2013). An example of this is the experiment conducted by Wimmer, Towsey, Planitz, et al. (2013); the authors compared traditional 20 minute manual avian point count surveys, at four sites over a five day consecutive period, to acoustic sensors. At each site, acoustic data was recorded with the sensors before, during, and after the manual surveys. The

acoustic data was analysed manually in a post-collection scenario to detect species that vocalised within. Approximately twice as many unique species were identified from the analysed acoustic sensor data, than identified in the field using traditional survey methods.

In bioacoustics scenarios, either ecologists or field workers are tasked with deploying sensors and then collecting the audio. Acoustic sensing is most often used for answering two basic ecological questions: species richness and presence/absence studies. Species richness studies measure the total distinct species captured by audio – this does not take into account abundance (Kindt & Coe, 2005). Alternatively, presence/absence (PA) studies measure the variation and abundance of species. Both methods require faunal vocalisations to be analysed but have different requirements for the volume of acoustic events that must be annotated.

This thesis focusses on terrestrial vocalising fauna. Marine bioacoustics is popular but also is significantly different enough in practice and method to be investigated separately. Vocalising fauna include any species that produce audible signals, including species from Insecta, Anura, Mammalia, and Aves.

2.2.1 Faunal Vocalisation Patterns

A vocalisation occurs when an animal emits some form of sound. There are many reasons for vocalising, including: communicating with a mate or flock; claiming or defending territory; as a response to a predator; begging for food; or commonly, attracting mates (Brown, Chaston, Cooney, Maddali, & Price, 2009). Emitting vocalisations is energy expensive for most fauna; hence, it is rarely done without reason (Slater, 2003).

Bioacoustics is not limited to any one family or species. However, Aves commonly exhibit rich differences in vocalisation between species. Because of this, they are often the subject of greater attention. There are some 10 000 species of birds known (James Clements, 2007; Gill & Wright, 2006). Only a fraction of those species are capable of vocalising (and thus being detected by acoustic sensors) and present within any sampled geographic area.

Acoustic vocalisation is an effective form of communication for birds, compared to other forms of communication like visual, chemical, or tactile (Catchpole & Slater, 2008). Acoustic communication allows large amounts of information to be conveyed within a short period; visual exposure is not necessary and acoustics are effective, both night and day. There are generally two types of bird vocalisation: calls and songs. Songs are more complex and longer than calls, whereas calls tend to be more functional, occurring year round.

Most Australian avian species call, rather than sing – they have simple call structures and relatively short vocalising periods with few syllables (Catchpole & Slater, 2008; Keast, 1993). For Australian fauna, forest birds tend to have low-frequency calls that attenuate less over distance (Keast, 1993). Keast (1993) also found that when vocalising maximally, avian species vocalised at rates of 20-30 seconds per minute.

For ease of description, the syllables that make up individual vocalisations have been broadly classified by researchers with descriptions of their visual appearance on a spectrogram (Duan et al., 2011; McCallum, 2010). The types of syllables include clicks (broadband events, a long vertical line); whistles (pure tones, horizontal lines); slurs (a tone that changes in pitch, a diagonal line); warbles (a tone modulated in one direction and then back, a 'V' or inverted 'V' shape); blocks (intense broadband events, darkened rectangles); oscillations (a series of oscillating clicks or slurs, repeated vertical lines); and static harmonics (stacked tones resonating from a fundamental frequency, parallel horizontal lines stacked above one another).

Example of all the aforementioned features have been included in *Appendix G – Examples of Faunal Vocalisations*.

2.2.2 Audubon

The [North American] National Audubon Society's Annual Christmas Bird Count is one of the oldest citizen science projects that exists today (National Audubon Society, 2010). This project has run a survey every year for 110 years that identifies what birds are spotted in particular regions. The survey runs for a week each year all over North America. This project is made possible by community involvement and has collected invaluable statistics on bird populations and migrations. While technically a bioacoustics project, all data is collected manually by participants.

2.2.3 eBird

eBird is a website dedicated to digitally collecting bird observations (B. L. Sullivan et al., 2009; Wood, Sullivan, Iliff, Fink, & Kelling, 2011). It was founded in 2002 by the Cornell Lab of Ornithology and the [North American] National Audubon Society. The website is a digital birding checklist that benefits from community interaction.

eBird is a hosted website with a large amount of active users. The records they collect are only sightings (including hearing a vocalisation). A community validates these records based on plausibility of the sighting and the reputation of the records' submitter. Importantly, unlike other related projects, eBird does not collect audio data. Thus, once their data has been collected there is no way to ground truth the data.

Despite the lack of raw data, relying on just the community's sighting records has been adequate for the project. The eBird project has published significant research and is considered successful in both the ornithological and eScience research disciplines. See http://ebird.org/content/ebird/about/publications/ for more information on their publications.

The Cornell Lab of Ornithology is known for publishing and maintaining the *Clements Checklist of Birds of the World* (JF Clements et al., 2012). This list is an invaluable resource to Ornithology experts worldwide.

2.2.4 WhaleFM

WhaleFM (Sayigh, Quick, Hastie, & Tyack, 2013) enables volunteers to classify recorded whale vocalisations, in order to promote research into the whales and the vocalisations they make. The site presents short segments of audio along with the spectrogram and asks the user to assign a single recording to a group with common acoustic elements. The project was founded by the Zooniverse organisation (the parent of Galaxy Zoo) and is another example of a bioacoustics project. What makes this project special is the use of sound and spectrograms, as the data needs to be classified at a mass scale. In its workflow, automated work is done to extract interesting parts of the audio files to show to participants – making this a semi-automated analysis process.

2.2.5 Xeno-canto

Xeno-canto is a website dedicated to hosting audio files of birds (Xeno-canto Foundation, 2013). The files are shared among a sizeable community of enthusiasts, and Creative Commons licences (simple, liberal, sharing licences) on uploaded audio are encouraged.

The audio recordings hosted by xeno-canto originate from their community – users are able to upload personal recordings. The recordings experience large variations in location and audio format. xeno-canto focusses only on avian vocalisations (birdcalls). Rather than being a sensor driven approach, recordings are typically short and generally feature one species at a time. Many of the recordings have been collected by participants with hand-held microphones, the result of which are recordings with good signal-to-noise ratios.

Community participation and collaboration is an important part of the website. Collaboration for the identification of unknown vocalisations functions well because of its active community. Part of the community classification process includes a set of rules for classifying the quality of audio recordings (Xeno-canto Foundation, 2012). The relevant excerpt is included below:

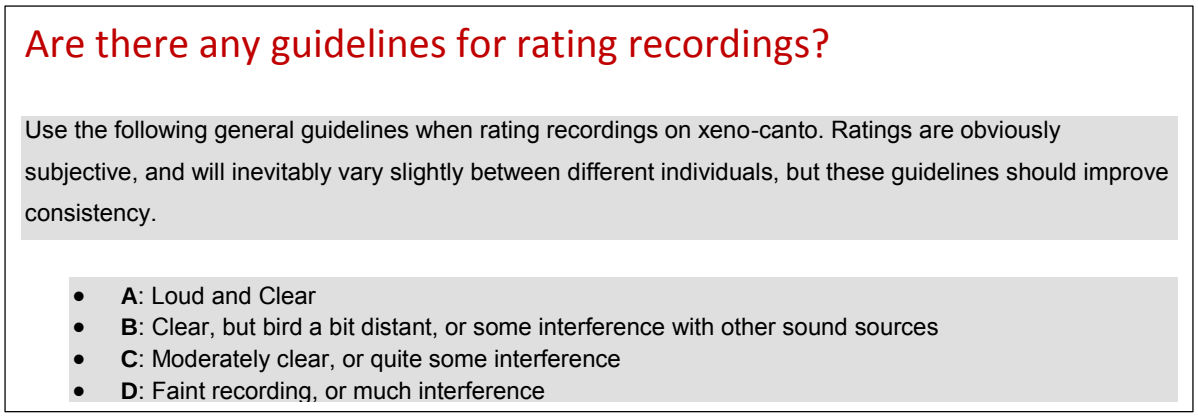


Figure 5 – An excerpt from the xeno-canto website

2.2.6 Pumilio

Pumilio (Villanueva-Rivera & Pijanowski, 2012) is a bioacoustics software package created at Purdue University. It is an open source project designed to archive, manage, analyse, and visualise audio produced by sensors. For an automated approach for detecting audio data, Pumilio includes sub-systems for running scripts for analyses in a parallel way.

Pumilio is not a hosted service but rather is a deployable software package that organisations can set up. This means a Pumilio instance can be installed on a private server, making the storage and analysis of sensitive data practical.

Pumilio has familiar concepts common to other bioacoustic software packages, like: recordings, tags/annotations for bioacoustic events, automated analysis jobs, and methods for organising audio recordings into logical containers like *Projects* and *Sites*. A screenshot (in Figure 6) demonstrates the software's ability to render spectrograms and playback audio.

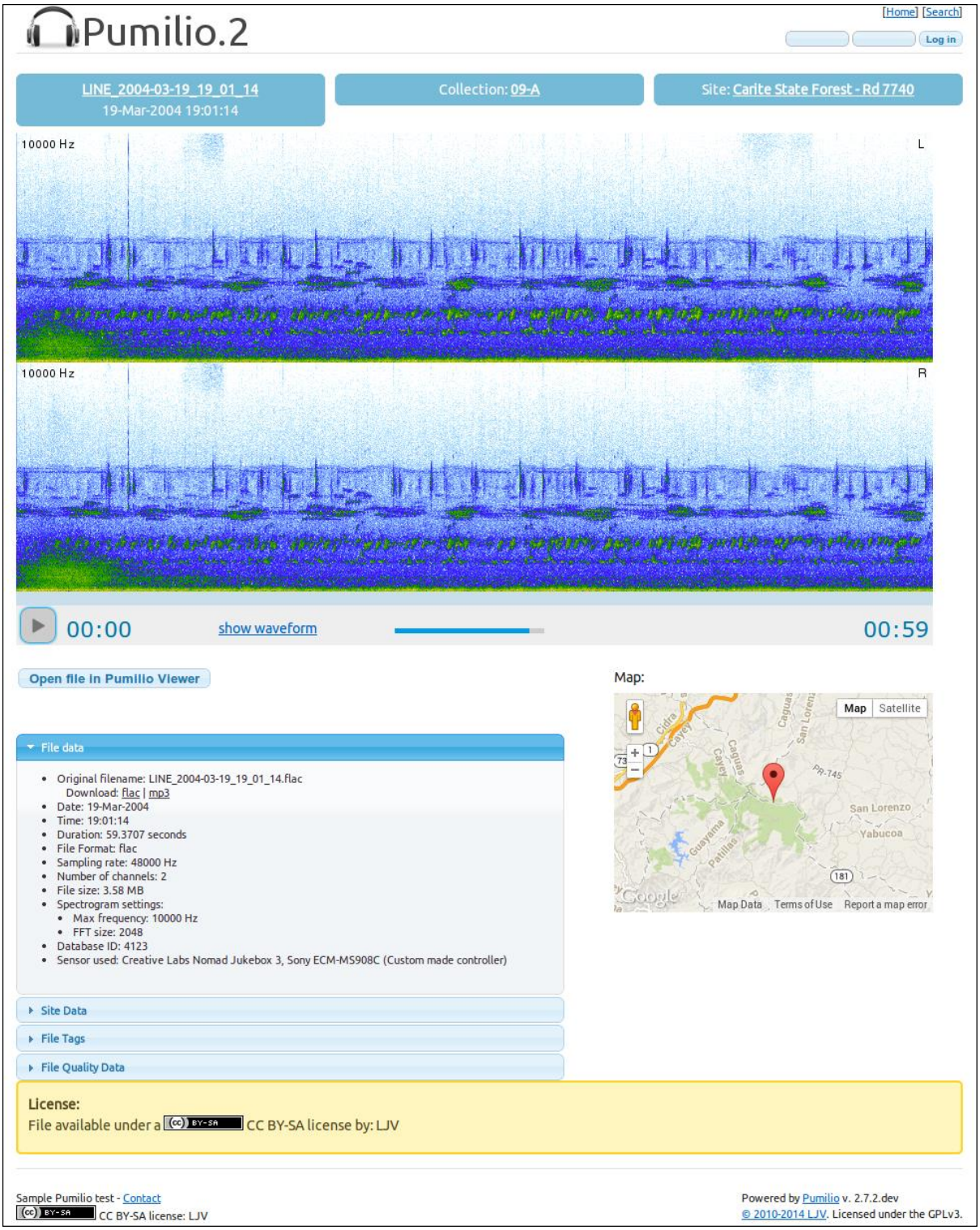


Figure 6 – A screenshot of the Pumilio Bioacoustics Software (http://ljvillanueva.github.io/pumilio/screenshots/)

2.2.7 ARBIMON

Automated Remote Biodiversity Monitoring Network (ARBIMON) is a commercialised bioacoustic software product from Puerto Rico (Aide et al., 2013). ARBIMON is a hosted website that charges approximately $0.15USD per audio upload.

Uniquely, part of their package is a fixed hardware design that is coupled with the software. The hardware records 1-minute recordings every 10 minutes and sends them back to a base station.

ARBIMON provide a species identification interface and an ROI (region of interest) detector interface as part of their modern, HTML5 based, software. It too has the standard features of other bioacoustics software packages (e.g. view, listen, annotate, and manage). The (human) interactive training of machine learning algorithms for species recognition is novel demonstrated potential.

2.2.8 Song Scope

Song Scope is a bioacoustics software package designed to detect acoustic events from audio gathered in the field (Wildlife Acoustics, 2011). It uses human input to define acoustic patterns of interest that can then be detected. Song Scope is an automated analysis system that requires an initial human configuration for each per species recogniser that is to be configured. The Song Scope's recogniser extracts MFCCs from audio recordings and classifies them with *Hidden Markov Models* (HMMs) (Agranat, 2009, 2013). Additionally, apart from the audio data and a human-entered template, the software uses only audio data to help predict faunal events.

Practically, the implementation of Song Scope limits its scale and flexibility. The program is limited to processing 2GB of audio data at a time, limiting it to a scale of a few days' worth of data. In terms of accuracy, Waddle et al. (2009) used the Song Scope product to detect bird calls; they found in a general setting the software detected false negatives in 45%-51% of the recordings tested. The authors of the paper advised caution when using Song Scope as the only method of detection (Brown et al., 2009; Duan et al., 2013). An overabundance of false positives is typically considered better than the alternative of potentially missing true positives.

2.2.9 Raven

Raven is a software package developed by the Bioacoustics Research Program at The Cornell Lab of Ornithology (Bioacoustics Research Program, 2011). The software package has modules for acquisition, visualisation, measurement, and the analysis of acoustic data. Raven is designed for detecting birdsong – particularly syllables of birdsong – and provides two basic analyses to the user. The first is an amplitude detector that detects regions of a signal where the magnitude exceeds a predetermined threshold. The second, labelled a Band Limited Energy Detector, uses a spectrogram;

it determines the level of background noise in the data and then selects regions that exceed a predetermined threshold in the relevant frequency bands.

Importantly, like Song Scope, Raven's detectors are trained manually per species that needs detecting. In terms of audio analysis, a batch mode makes scaling data analysis more efficient and several training models can be detected at once. However, since Raven needs each training model to be trained per species by humans, it suffers the same scale problems as Song Scope (Duan et al., 2013). Raven is not designed for faunal event classification; rather, it is better suited for event detection in large volumes of data, rather than for general automated analysis.

2.3 Bioacoustic Data Collection

A survey of the literature reveals that existing methods of data collection are varied. Bioacoustic data can be collected informally or formally and indirectly or directly. This section details popular terrestrial collection formats.

2.3.1 Manual Recreational

There are people that enjoy nature and regularly seek out fauna as a recreational pastime; this audience, especially for avian species, is catered to by many websites. Two example of websites dedicated to *birders* (recreational avian seeking enthusiasts) are Xeno-canto for example and eBird. Any data that is gathered is often small and for personal use. Checklists for sightings or acoustic events are popular and the most common. Recording birdcalls for personal collections is becoming more popular; these recordings are typically short and focussed on one species (unlike unfocussed and voluminous sensor data).

Field guides (traditionally in book form, increasingly in digital form) like *The Princeton field guide to the birds of Australia* are often carried to assist identification (Simpson & Day, 1996). Carrying a physical field guide has evolved into carrying a guide on a laptop and recently, into carrying a guide on a smart phone. Often these digital guides also include recorded examples of species vocalisation – a powerful innovation. However, these digital audio guides are often produced in an ideal environment and may not be accurate representations of real world audio, because background noise and acoustic soundscapes have a significant effect on the way audio is perceived (M. W. Towsey & Planitz, 2010).

2.3.2 Manual Surveys

Manual surveys of areas are common formal methods for monitoring biodiversity. They are well-defined practices undertaken by recreational participants or scientists that need data for a region. An example of a survey from the Atlas of Living Australia is included in *Appendix E – ATLAS Record*

Form. Surveys can record various ecological indicators (other than acoustic events) to assess the environment.

It is widely recognised that conducting this work manually is difficult and hard to standardise – especially across large geographical regions (Waddle et al., 2009). The book *Systematics and Taxonomy of Australian Birds* (Christidis, Boles, & Ornithologists' Union, 1994) is an exceptional resource for understanding Australian avian fauna though manual surveys.

Morphology is the study of shape, size, structure, and appearance. Traditional classification techniques are based on morphological traits (Acevedo, Corrada-Bravo, Corrada-Bravo, Villanueva-Rivera, & Aide, 2009). When a human is conducting a field study, attributes such as shape, size, structure, and appearance to are used to visually differentiate between species. However, this can prove ineffective because these traits usually need trained participants to be involved.

2.3.3 Acoustic sensors

Acoustic sensors are an alternative to manual recreational or survey approaches. Acoustic sensors are designed to collect voluminous, unfocussed, audio data from the environment in which they are deployed. Acoustic sensors can be used to monitor terrestrial or marine fauna; however, the approaches, methodology, and equipment required vary considerably (Rickwood & Taylor, 2008).

Using sensors to record audio is superior to manual surveys as data can be collected objectively, continuously, and without the bias of human presence. It is unfeasible and expensive for a field worker to compete with the spatiotemporal efficiency of sensors. When a field worker does a survey, the original data – what they hear and see – is not kept. Conversely, a sensor's data is recorded and that digital record can be referred to again whenever necessary. The ability to recheck the source data is called ground truthing and is a principal advantage of permanently storing the data collected by sensors. There is evidence that acoustic sensors can work as well or better then manual surveys. (Holmes, McIlwrick, & Venier, 2014; Wimmer, Towsey, Roe, et al., 2013).

There are many examples of bioacoustic research that rely on sensors, including: Agranat (2009, 2013); Aide et al. (2013); Bardeli et al. (2010); S. Brandes (2008); Butler et al. (2007); Frommolt, Tauchert, and Koch (2008); Gage, Napoletano, and Cooper (2001); A. N. G. Kirschel et al. (2009); Mason et al. (2008); McIlraith and Card (1997); Riede (1993); Sayigh et al. (2013); Somervuo, Harma, and Fagerlund (2006); Taylor, Watson, Grigg, and McCallum (1996); Tucker et al. (2014)

2.4 Bioacoustic Sensor Data Analysis

Acoustic sensor data must be analysed after it has been collected. Unlike short, focussed, and curated recordings that focus on one call (commonly recorded by ecologists, biologists, or

recreationalists), sensor data is voluminous and contains much irrelevant data. Typically, analysis is conducted in one of two ways: by extracting representative summary statistics or by detecting and classifying individual faunal vocalisations. This section will explain both techniques as well as the difficulties involved in conducting analysis.

Analysis can be conducted with human analysts or by automated algorithms. Most existing research utilises automated algorithms, however, this section will demonstrate that despite advances in automated detection of fauna in acoustic sensor data, a complete solution is still a significant challenge. Data management and creating algorithms that can scale for months of acoustic sensors data are difficult research problems (Planitz, Roe, Sumitomo, Towsey, Williamson, & Wimmer, 2009).

2.4.1 Difficulties Analysing Bioacoustic Data

Analysis techniques suffer when processing poor quality audio data. Noise, range, and geographical variance in acoustic signals can require that some data be discarded, as if it were never collected. Following is a summary of the problems encountered while analysing acoustic sensor data for faunal events. These problems are experienced by both human analysts and automatic forms of analysis, though to different effect. Each of the following problems are common for most acoustic sensor data collected.

2.4.1.1 Noise

The signal in sensor audio data contains various types of noise. Noise is simply defined as *"unwanted signal that interferes with the communication or measurement of another signal"* and noise itself is still signal encoded with information (Vaseghi, 2008).

Noise sources can stem from the sensor itself (due to faulty microphones or sensor circuitry), natural sources, or manmade (*anthropogenic*) sources. As the definition of noise depends on context, noise can be signal generated by unimportant sources. For example, traffic (cars / trucks / aircraft / boats), humans, machinery, and others sources all generate audio signal that can interfere with, overlap, or disrupt the native acoustic soundscape (S. Brandes, 2008; Hu et al., 2009; M. W. Towsey & Planitz, 2010).

Moderate wind and rain can dominate the signal of an acoustic recording (Aide et al., 2013). Additionally, for most fauna the regularity and type of vocalisations emitted correlate to environmental events like wind and rain (A. N. G. Kirschel et al., 2009; Taylor et al., 1996). Lastly, it is common for several species to vocalise simultaneously. This results in acoustic events where the audio signals are intertwined. On a spectrogram, it is possible to see the components of the acoustic events from two (or more) sources overlapping. Extreme examples include species that vocalise so

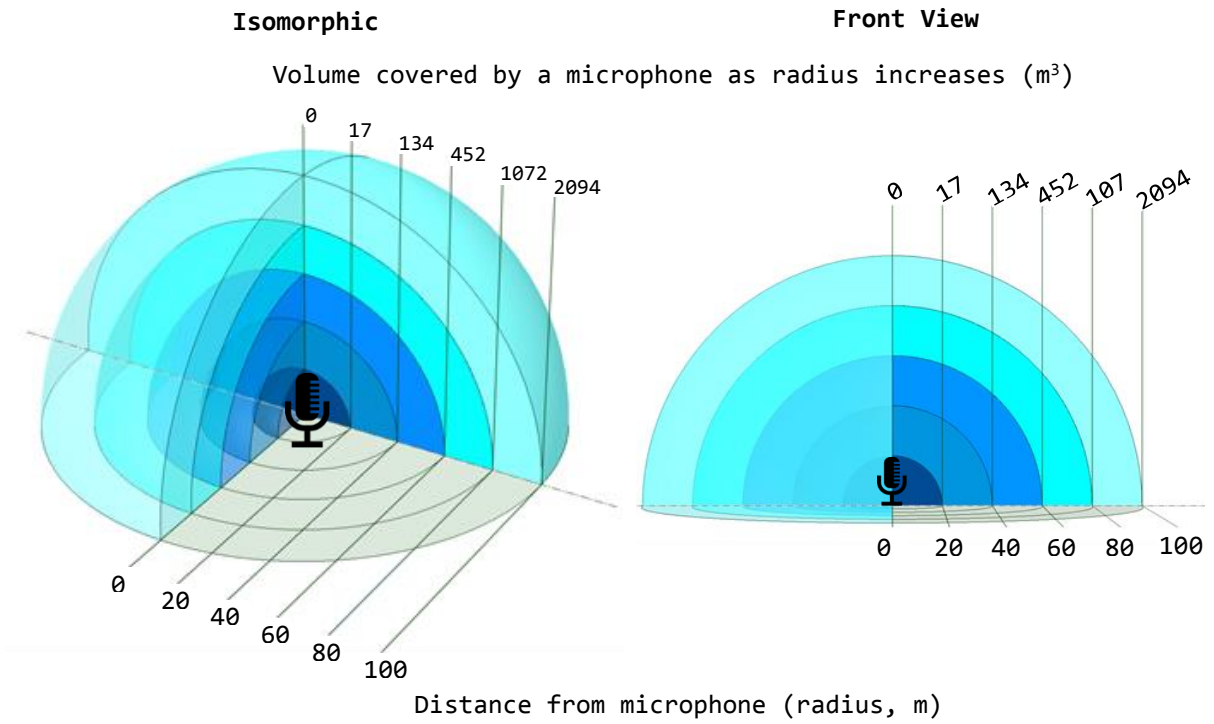
intensely or consistently, that they dominate entire portions of an acoustic space. Crickets and cicadas are examples of dominating signal sources.

2.4.1.2 Range

Range plays a large part in the quality of an acoustic event. The propagation of sound follows an *inverse square law*. This law states that the amplitude of a signal is inversely proportional to the distance travelled. This attenuation of a signal is an observable form of the law of conservation of energy.

Because the microphones used by most sensors are (nearly) omnidirectional, acoustic events are detected from a sphere around the microphone – where the radius is a function of the sensitivity of the microphone. The sphere is not complete and is usually hemispherical because hard surfaces dissect it, such as the local terrain (like flat ground, a rising slope, or some other topology).

Regardless, the sensing area can be divided into a series of inner spheres at regular intervals inside the main sphere, where the main sphere's radius is determined to be the point where acoustic signal can no longer be detected by the microphone. Moving from inner spheres to outer spheres results in an increase of the radius from the central point (the distance from the microphone). The volume of each successive shell (or hollow sphere) is proportionally larger than the radius. In addition, as the radius increases, a signal of the same amplitude is perceived as quieter (as per the inverse square law) by the microphone. This relationship is shown in Figure 7.



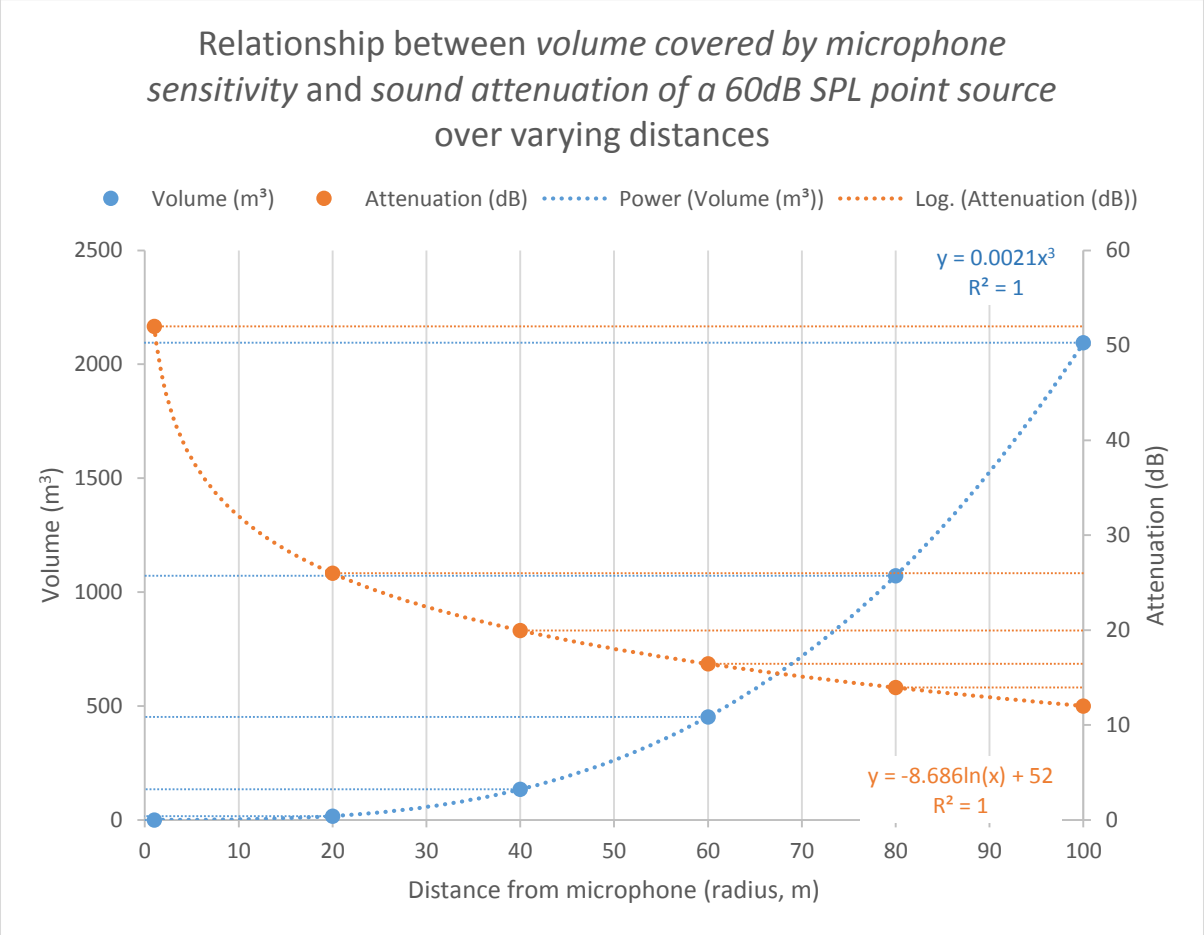


Figure 7 – A generalised example of the relationship between sound attenuation and a microphone's sensitivity. Assumptions made include an even distribution of sound sources (e.g. vocalising fauna); consistent atmospheric conditions; a flat and constant geological topology; perfectly omnidirectional microphones; and consistent vegetation.

Using this reasoning, as a probability, acoustic sensors are much more likely to detect quiet acoustic events, as opposed to acoustic events that are closer (and louder) to the sensor. In other words, clear or loud acoustic events are probabilistically uncommon.

The attenuation (damping) a sound wave experiences varies by the frequency of the sound wave (DIN ISO 9613-1, 1993). For two sounds of equal amplitude, heard from the same distance, where one has a higher frequency then the other, the former will experience more attenuation than the latter. The amount of attenuation depends on several factors including the frequency emitted, atmospheric temperature, pressure, and humidity. The topology and vegetation of an environment also affects sound damping.

Thus, given two identical vocalisations, at different distances from the observer, the vocalisations will sound and look (on a spectrogram) different from each other. Torresian Crows (*Corvus orru*) produce vocalisations that demonstrate this behaviour well. Typically, the vocalisations are short broadband "barks". However, the further away the crow vocalisation is from the sensor, the more the higher frequencies of the vocalisation experience attenuation. Figure 8 shows an example of crow calls from separate individuals calling within the same minute. The individuals are at different distances away from the sensor. Notice the reduced height (high frequency portion) of the vocalisations in the middle of the example; these are the reply calls from a crow that is further away from the sensor.

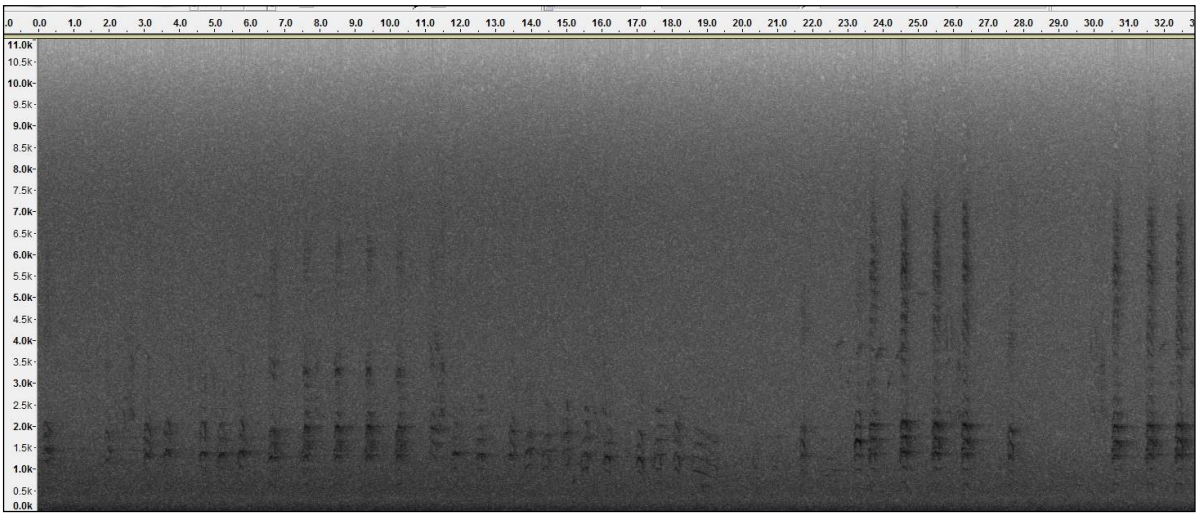


Figure 8 – A spectrogram showing the effect of sound damping for three Torresian Crows. The individuals are at varying distances from the sensor. Audio sample: SERF Acoustic Study, South East site, 16th Oct 2010, 06:37

2.4.1.3 The Variance of Vocalisations for Species over Regions

It is well known in ecology that the same species, in separate geographies, can have different vocalisations (Catchpole & Slater, 2008; Kroodsma & Miller, 1996; Marler & Slabbekoorn, 2004, p151). Most times these variations are small but they can be large enough to interfere with

automatic detectors and in some cases, human analysts as well. This means a one-size-fits-all approach cannot be applied to a single species without catering for region-based variance in vocalisations.

2.4.2 The Role of Human Speech Recognition

For general acoustic problems and human speech pattern recognition in particular, large bodies of research have already been conducted. For human speech recognition, both signal detection and classification are relatively advanced (Acevedo et al., 2009). Ongoing research in these areas, including research into more capable classification techniques continues, but complications like accents and speech impediments still pose a major challenge (Scharenborg, 2007).

In the context of bioacoustics, many techniques for human speech recognition are not compatible with faunal acoustic event detection, especially for real-world data (Planitz, Roe, Sumitomo, Towsey, Williamson, Wimmer, et al., 2009; M. Towsey, Planitz, Nantes, Wimmer, & Roe, 2012).

Human speech occurs in a relatively narrow set of frequencies, with a limited set of producible sounds. Human speech follows predictable patterns: the entire set of syllables, the bandwidth of speech, the cadence, and various other features are all reasonably consistent (Doupe & Kuhl, 1999).

For the task of recognising fauna, all of the aforementioned features can vary considerably between species (Catchpole & Slater, 2008; Härmä, 2003). This variation is good – it allows species to be distinguished easier – however it means that assumptions made for human speech recognition techniques are violated.

For example, MFCCs (section 2.1.4.3) are commonly used in human speech recognition to detect static harmonics. However, static harmonics are far less common for faunal vocalisations, making MFCCs less practical. Additionally, birds vocalise in frequencies and syllables that are different from human speech. Some bats vocalise in ultrasonic frequencies whilst other animals (like some frogs and some whales) vocalise at very low frequencies.

Most speech recognition algorithms assume a relatively noise free environment, with a high SNR, and repeatable signal capture. This is not comparable with the audio data that sensors collect – non-repeatable events in noisy, low SNR ratio audio. A contributor to SNR is distance of the sound source from the microphone. Speech recognition algorithms assume the signal source is close to the microphone, whereas the opposite is generally true for sensor-based approaches (see 2.4.1.2).

Despite the differences between human speech recognition and faunal classification, some techniques are still applicable. Many ML algorithms can classify unique data instances and either classify them as an existing class or as a new class. This concept is important in human speech

recognition in detecting accents, accounting for noise, or mispronunciation of words (Skowronski & Harris, 2006). ML could offer the same advantage for acoustic faunal classification – accounting for noise or regional variation in species (inclusive classification) and separating out never-before-heard species (exclusive classification).

2.4.3 Automated Methods of Bioacoustic Event Recognition

Successful automatic detection algorithms occur when trying to detect one particular species that has a distinct, easily identifiable vocalisation. For example crickets (S. Brandes, 2008), grasshoppers (E. D. Chesmore & Ohya, 2004), and cane toads (Hu et al., 2009), are all considered easy to detect, due to the nature of their consistent and repetitive vocalisations.

Apart from non-specific algorithmic techniques for detecting acoustic events in audio data, Machine Learning (ML) has been used for faunal acoustic event detection for various research projects (Bardeli, 2009; M. W. Towsey, Wimmer, Williamson, Roe, & Grace, 2012). Machine Leaning is commonly used for call classification of faunal acoustic events rather than for signal detection (Acevedo et al., 2009). This is reasonable as many ML algorithms are designed to be classifiers – to match one set of variables to another. Signal processing is a different task and is treated as such.

There are many examples of the use of automated algorithms (and ML) for the detection of faunal acoustic events. A few examples are listed below:

- Frogs (T. S. Brandes, Naskrecki, & Figueroa, 2006; Han, Muniandy, & Dayou, 2011; Taylor et al., 1996; Waddle et al., 2009)
- Cane toads (Hu et al., 2009)
- Birds and Amphibians (Acevedo et al., 2009; S. Brandes, 2008; Härmä, 2003; Lazarevic, Harrison, Southee, Wade, & Osmond, 2008; Mason et al., 2008; Planitz, Roe, Sumitomo, Towsey, Williamson, & Wimmer, 2009; Planitz, Roe, Sumitomo, Towsey, Williamson, Wimmer, et al., 2009)
- Bats (Herr, Klomp, & Atkinson, 1997; Skowronski & Harris, 2006)
- Marine Life (presence / absence detection) (Moore, Stafford, Mellinger, & Hildebrand, 2006; Palialexis, Georgakarakos, Karakassis, Lika, & Valavanis, 2011; Sayigh et al., 2013)

The previous references are a small sample of automated acoustic event detection research. There are varied levels of success in the approaches of the research mentioned. However, what this sample demonstrates is that the research currently being conducted involves techniques that are limited to species, regions, or vocalisation type.

Many automated event detection research projects make concessions with their scope. For example, often the audio events that need to be matched are segmented from the audio data manually before automated classification occurs. Two examples of this can be found in Acevedo et al. (2009) and Skowronski and Harris (2006).

Acevedo et al. (2009) conducted a study on avian and anuran vocalisations to determine what type of classifier worked best. Importantly, the vocalisations were labelled by hand, as it was outside the research's scope to segment calls automatically. Additionally, they tested two forms of feature extraction: a simple bounds approach (bandwidth, energy, and duration) and a more complex 11-stage energy profile extraction. The complex feature set performed better, however the simple features had an acceptable performance. Their experiment was conducted on a small dataset.

Skowronski and Harris (2006) used human speech recognition algorithms to detect *microchiroptera* (micro echolocating bats). Again, vocalisations were labelled manually as it was easier than automatically segmenting the calls. The authors also used "global features" to represent the calls – these features included duration, bandwidth, and other meta-data. The results of the paper found that generally more features performed better, though overlapping calls were difficult to distinguish.

Developing only species-specific recognition techniques is not feasible and fragments research efforts; thus, research into generalised approaches for automatic detection of faunal events is justifiable. Examples of generalised recognisers can be seen in the research of Duan, Zhang, Roe, Towsey, and Buckingham (2012).

To summarise, for automated bioacoustic event detection much research has been done. However, much of this research is either species specific, constrained to smaller studies, or only addresses parts of the problem (i.e. classification over event detection). Additionally, there is precedence for simplistic bounding-box style feature sets that describe audio events by their dimensions.

2.4.4 Acoustic Ecological Indexes

As an alternative to detecting and classifying faunal acoustic events, research has been conducted into calculating indexes that link acoustic soundscapes to ecological conditions (Gage et al., 2001; Sueur, Pavoine, Hamerlynck, & Duvail, 2008; M. Towsey, Parsons, & Sueur). Indices are an abstract way of evaluating biodiversity that focus on the correlation between extractable large-scale features of audio and faunal activity. Indices are typically defined over large amounts of data (e.g. one value per minute) allowing for the large-scale visualisation of data (M. Towsey, Zhang, et al., 2014).

Various properties of audio can be used as indices, such as average amplitude, spectral or temporal entropy, or SNR (Sueur et al., 2008; M. W. Towsey et al., 2012). Acoustic indices specifically designed

for bioacoustics have also been developed. The Acoustic Complexity Index (ACI) quantifiably measures the change in biotic vocalisations, whilst filtering out anthropogenic noise (Pieretti, Farina, & Morri, 2011).

Depraetere et al. (2012) designed an acoustic richness index based off a combination of temporal entropy and the median of the amplitude envelope that filters out anthropogenic noise. Their study found a correlation between their indices and ecological condition. Gasc et al. (2013) conducted a similar study to measure the correlation of acoustic indices to phylogenetic differences. This study was conducted over a large dataset (19 420 sites). They concluded that spectral indices had the highest correlation.

Although showing promise, acoustic ecological indices do not directly address the species richness ecological question that this research does. The information they produce is useful in large-scale contexts but is unable to provide details for vocalisations of particular species. However, it is possible to do the reverse, to transform annotated faunal events into indices that can be used with other indices, indicating that ecological acoustic indices are a valuable way to visualise and summarise event data.

2.5 Semi-Automated Annotation of Bioacoustic Vocalisations

Semi-automated annotation of bioacoustic vocalisations is a broad term that includes all analysis techniques that involve significant human and algorithmic components (D. Chesmore, 2007). Such techniques utilise the complementary strengths of human and algorithmic analysis techniques (discussed in2.1.5). For bioacoustics, many of the parent ecological projects have unused human resources (for example citizen scientists) – using those people as analysts in a formal analysis system has potential.

Solving the fully automated analysis task (automatic classification and identification of faunal acoustic events) is complex and accurate solutions take time to develop. The previous section demonstrated that most existing automated techniques attempt to solve only parts of a problem. If they were intended as such, they could be classified as semi-automated methodologies since significant human effort is involved in conducting their analysis. Additionally, the results of most automated methods still need to be verified by a human analyst; human effort is always required (A. N. Kirschel et al., 2009).

Semi-automated analysis is conducted because humans are better than algorithms for some tasks. Semi-automated analysis should produce analysed data faster than manual methods but will be much slower than fully automated processes. Yet producing *medium velocity* data is still valuable.

Following are some examples of related semi-automated research loosely categorised (not mutually exclusive) based on the value they extract from analysts.

2.5.1 Data Scale Reduction

Using a semi-automated analysis method to reduce the amount of raw data participants need to analyse can increase the efficiency significantly. An example of a simple form of such a system can be found in the work of Wimmer, Towsey, Roe, et al. (2013). They statistically determined the periods of the day where the most vocalisation diversity was found (the three hours around dawn). Those samples are then provided to human analysts, resulting in the most unique species analysed for the least human effort.

Acoustic indexes have been used in semi-automated methodologies. M. Towsey, Wimmer, Williamson, and Roe (2014) use acoustic indices to *smart sample* the most relevant minutes of audio with a large body of audio. The most diverse samples are sent to human analysts that segment and classify acoustic events. The result is a higher diversity of species classified with less human work.

There is precedence for using advanced algorithms. Tachibana, Oosugi, and Okanoya (2014) use SVMs to reduce the data that human analysts need to analyse. Analysts still perform classification, but song and syllable detection is done by the algorithm. Separately, Potamitis, Ntalampiras, Jahn, and Riede (2014) conducted research that automated the analysis of audio data to detect avian species. Real world, long duration acoustic sensor data was processed with *Hidden Markov Models* (HMMs). Their experimental results for detecting two avian species were promising but yielded a large number of false positives. They conclude that automatic species recognition reduced the space (size of data) required for a human to analyse substantially but human analysis was still required: "*technology is not yet mature enough to completely automate decision making on critical bio-diversity issues*". The authors conclude that taxon-specific detectors are necessary for the best results with fully automated analysis methodology. Yet there is still potential to further their semi-automated methodologies.

2.5.2 Classification

The Zooniverse organisation runs several projects that involve semi-automated analysis; of note are Galaxy Zoo and WhaleFM. Galaxy Zoo (previously discussed in section 2.1.2) uses citizen scientists to classify the morphologies of galaxies. Importantly, the segmentation of the galaxies (into smaller images from much larger images) was done automatically as part of the Sloan Digital Sky Survey (Lintott et al., 2008). The combination is a proven example of the effectiveness of semi-automated analysis methodologies. For WhaleFM (previously discussed in section 2.2.4), a large variation in the call types of killer and pilot whales, made purely automatic recognition difficult (Shamir et al., 2014).

The WhaleFM project combined computer analysis methods with citizen scientists to analyse their data, thus creating a semi-automated system. The citizen scientists were better at separating the vocalisations based on the geographical region in which they were found but automated methods proved more accurate.

Kubat, DeCamp, Roy, and Roy (2007) have conducted research into transcribing speech and determining the head direction of a newborn baby. With multiple cameras and microphones installed, the project recorded data for 22 months. When trying to analyse the data, the authors found that traditional human speech recognition techniques were inadequate for several reasons. To analyse the data anyway, the authors switched to a semi-automated approach for conducting their analysis. Algorithms were used to detect possibly relevant sound bites amongst the masses of audio data and human users were used to transcribe the audio. Their tool also attempted to automatically detect face direction to save time but allowed users to correct the data if it was wrong.

2.5.3 Bootstrapping

Bootstrapping an automated analysis refers to a process, usually manual, that sets up or prepares the analysis. Using human analysts to segment, label (classify), or otherwise process data is a valid form of bootstrapping. Similarly, both Song Scope and Raven (see sections 2.2.8 & 2.2.9) use human analysts to define a query to search for within larger pieces of audio. Typically, though, bootstrapping refers to larger datasets. One result of bootstrapping data is that larger amounts of training data are available for fully automated analysis techniques (Wolf, 2009). Additionally, initial ecological inferences can be made from bootstrapped data, allowing for future flexibility in analysis.

S. Brandes (2008) identified the advantages of semi-automated analysis stating that a library (where the library is a set of bootstrapped data) of example faunal vocalisation is important for most detection systems. Libraries that capture the species in their native environment, with secondary calls and responses, are considered particularly useful (S. Brandes, 2008). For accurate, real world, detection tasks the creation and population of a library of reference calls is needed, both for reference by human participants and as training data for ML algorithms. Brandes (2008) hypothesises that using semi-supervised learning – the population of a reference library by supervised human participants – will allow software classifiers to be trained accurately and thus improve overall performance.

Tachibana et al. (2014) use humans to train SVM classifiers for detecting birdsong. Their manual analysis process did not scale for large amounts of data. Their research focussed on reducing human work by using better machine learning to minimise the number of instances a human has to analyse.

This research used human analysts to bootstrap their automated algorithms by analysing an initial data set. Their experiment measured accuracy of the automatic classifier once bootstrapped on larger datasets. Their work demonstrates the semi-automated methodology well.

2.6 Labelling Acoustic Events

Professionally encoded metadata is considered high quality but does not scale due to the lack of experts (Mathes, 2004). These experts are organisation employees, or professionals, that are too busy or cost too much to employ for the curation of data. The solution to this problem is to allow non-experts to encode metadata.

This section introduces the concepts of tagging and folksonomies and elucidates the advantages of using these systems to encode metadata for acoustic events.

2.6.1 Tagging

Tags are textual labels that are associated with a particular piece of data. Tags are frequently used to associate a resource with additional semantics. The Web 2.0 phenomenon has made tagging an increasingly popular Internet principal for classifying the data of many websites (Cuff, Hansen, & Kang, 2008). Tagging systems have become popular because they have a low barrier of entry for users (Gasc et al., 2013). This is a central reason for the proliferation of tagging throughout many popular websites, including Del.ico.us, Flicker, Twitter, Gmail, Facebook, YouTube, SoundCloud, and Tumblr. The user that tags content can be the author/owner or another member of the community – tagging can be personal or social.

Content can be easily labelled with tags because of their small, short, and descriptive nature. They are generally short and usually adjectives but can sometimes also be short phrases. Tags should not contain most of the unnecessary language that would exist in an equivalent descriptive sentence (Marlow, Naaman, Boyd, & Davis, 2006; Xu, Fu, Mao, & Su, 2006). This lack of function words required for grammatical correctness results in tags being linguistically simple enough to be easily used for comparison, categorisation, and classification purposes with relative ease by both humans and algorithms.

A tag is intended to describe one unique concept; tags should not combine atomic pieces of information (Xu et al., 2006). If a tag encodes multiple pieces of information, it loses its ability to describe a single concept uniquely, thereby resulting in less effective categorizations of content. To express multiple concepts or classifications, most tagging systems allow the association of more than one tag to a piece of content.

2.6.2 Folksonomies

Tagging systems are also interesting because they are often extensible by the community that uses them – a fixed set of tags (a taxonomy) can be enforced in a tagging system but usually is not. The term used to describe community created tags (classification schemes) is *folksonomy*. A *folksonomy* is defined as a 'folk' (community) generated taxonomy (Vander Wal, 2007). Other terms have been used to describe folksonomies such as '*collaborative tagging*', '*social classification*', '*social indexing*', and '*social tagging*'.

The result of allowing the 'free tagging' of data (information or objects) is that domain members can use their own vocabulary to describe the data. Herein lies the value of a folksonomy: the resulting specialised nature of the tags produced. Using domain members to describe data, produces a set of tags that consist of a domain-appropriate, specialised lexicon. In this sense, a domain member is usually a member of the community using the tagging system – they are experts at describing their own data (Marlow et al., 2006).

2.6.3 Linking tags

Tags generally have no interdependent structure. The concept they represent should be independent and association between tags is determined by their common placements on certain types of data.

An evolution of independent tagging systems allows for the definition of hierarchical tag structures that casually create direct associations between different tags. There are examples of tagging systems that infer hierarchical relationships (Heymann & Garcia-Molina, 2006). A folksonomy, like the more generalised notion of tagging, was originally defined as having no hierarchical structure (Mathes, 2004); this means the set of classes are simply a group of labels commonly used to describe a set of objects but they do not exist in a predefined or static structure. However, there have been examples of structured folksonomies (Marlow et al., 2006); for example, Gmail can create nested labels to organise emails into a folder-like hierarchy.

Excluding systems where explicitly creating relationships between tags are possible, there is value in inferring the relationships between the tags in other systems. Often the best way of forming relationships between tags is to do post-hoc analysis on the tags to determine the frequency of tags occurring on the same object. There are limitations to this method, including ambiguity (lack of context), erroneous associations, and the limited capability of algorithms to correctly identify synonyms. Meanwhile, the relative strengths include the ability to "browse", a user vocabulary that labels user content, a low barrier to entry, and a low cognitive cost.

Echarte, Astrain, Córdoba, and Villadangos (2008) described efforts to regroup syntactic variations of folksonomic tags with basic pattern matching techniques. The frequent multiple syntactic variations in tags for tag systems they surveyed were difficult to reconcile. The use of the Levenshtein and hamming metrics can be used for informal string matching to identify syntactic variation. While conducting their experiments they highlight and important disadvantage of using folksonomies: synonyms become prevalent when the folksonomy gets larger; typographical misspelling; grammatical; plural/singular versions of the same word; and even combinations of error types. The authors have had reasonable success using both similarity metrics in their experiments, but the Levenshtein algorithm performed better for cases that had missing or extra characters. Both of the techniques tested do not perform well for resolving singular/plural cases. This research by these authors has many parallels to the problems seen in bioacoustic folksonomic tagging systems. However, there is an important difference: many of the tags in normal folksonomies are adjectives, whereas many of the bioacoustic tags are nouns.

2.6.4 Annotating multimedia data with tags

There are many examples of multimedia annotation that uses tags: Flickr annotates images, SoundCloud annotates sound, YouTube annotates video formats, and Vannotea (Schroeter, Hunter, Guerin, Khan, & Henderson, 2006) can annotate other multimedia formats. Tagging has been used to annotate audio data for ecological science projects. Similar research projects have cultivated libraries of audio recordings that have been labelled – commonly these libraries capture short and focused recordings containg only the species of interest within the recording and thus often label entire audio recordings. The “Jacques Vielliard” dataset maintained by UNICAMP (Cugler, Medeiros, & Toledo, 2011) and the Berlin Sound Archive (Bardeli, 2009) are two examples. These libraries are excellent resources; however, the majority of their recordings are not collected from sensors. This has an effect on what analysis techniques are effective. Typically, human analysts (as opposed to automated analysis) can distinguish greater amounts of detail from acoustic sensor recordings.

2.7 Summary and Implications

This literature review covered a wide range of topics to demonstrate a gap in research that can be addressed by this thesis. Reviewing existing literature allows for efficient research, that leverages prior knowledge.

The literature revealed that there is a long tradition of wildlife monitoring. One of the greatest examples of this is Audubon project; they use community volunteers to gather data on Aves and have done so for over one hundred years. However, the majority of manual surveys are conducted by ecologists. As technology improves, it is becoming increasingly possible to scale wildlife

monitoring. This thesis focusses on monitoring terrestrial fauna with acoustic sensors. The literature demonstrates that various research projects around the world are using acoustic sensor recordings to monitor the environment. The fields of bioacoustics and ecoacoustics are well established.

Analysis of acoustic sensor data is required to distil massive amounts of raw data down into results that can be used to form ecological conclusions. The literature briefly touched on acoustic indices that are used for evaluating the acoustic soundscape. Although indices based analysis is becoming more popular, most existing research relies on detecting faunal acoustic events.

Detecting acoustic events in audio data is ideally an automated process. Automated methods are intensely researched but the literature showed that there is currently no single, high-accuracy, generalised, recogniser capable of detecting all faunal acoustic events in acoustic sensor data. Importantly it is evident that there is much research still being conducted into automated methods of detection; this indicates that the field has not yet reached its full potential.

While automated analyses remain an active area research, analysing data via other methods is still important. In particular, the literature shows that the semi-automated analyses of audio data can efficiently produce data and support ecologists. Ecologists can still use the smaller amounts of data that are produced; additionally, analysis produces data for researchers who can then utilise it as more training data.

The literature elucidated the strengths and weaknesses of human analysts. Human analysts are good at finding similarity or discerning differences in visual and audio data. However, they tire and lose interest, as well as struggle with the ability to memorise large amounts of data. With training and practice, experts are capable of memorising significant datasets; however, experts are still fallible and often are only considered experts for limited geographical regions.

The literature shows that there are software packages designed to assist the identification of animal vocalisations. Xeno-canto is an excellent web site that couples manual analysis and community sharing of acoustic data – however, it is designed for short, focussed recordings. Other software packages, like RAVEN and Song Scope, implement generic vocalisation detectors – however the reusability of this software was seen to be limited. The literature shows three major pieces of software designed for managing and analysing environment acoustic sensor recordings: Pumilio, ARBIMON, and the software produced by the QUT Ecoacoustics research group. Each of these software packages have similar aims and a focus on automated analysis. However, each could improve their user interfaces for the manual analysis of acoustic events.

Chapter 3
Background and Methodology

This chapter will describe the specific context in which this research was conducted. This thesis is concerned with assisting human analysts that annotate acoustic events and accordingly is best tested within an existing bioacoustics research project. This research was conducted for the QUT Ecoacoustics Research Group. This chapter details the shared background and methodology used for collecting and analysing data.

This chapter is written by monograph and additionally has a publication – authored primarily by the thesis author – associated with it. The majority of this chapter uses a monograph format to detail the methodologies and software artefacts that were used for this research. However, in the publication, additional, peer reviewed, published, material on the methods used can be found. As this is the methodology chapter and because the publication contains applied research, it is not considered part of the main narrative of this thesis. The publication is included verbatim at the end of this chapter (page 67).

The methodologies and software artefacts discussed within have been the subject of research for the author and other researchers (Cottman-Fields, Truskinger, Wimmer, & Roe, 2011; Mason et al., 2008; Wimmer, Towsey, Planitz, et al., 2013; Zhang et al., 2013). The author was also employed as a research assistant (software developer) for the research group.

This chapter will outline the methods used to collect data and the software created to expose that audio data to users on a website. A section on the different forms of analysis possible (automated / semi-automated / manual) is followed by a detailed description of how annotations are created. Annotations are the time and frequency bounded classifications of audio events; the process for creating them is core to this thesis's research questions. Lastly, the scale and types of data that were available for use in the thesis are described.

3.1 Data Collection

The QUT Ecoacoustics research group has been collecting audio data since 2007, using a variety of different sensors deployed to different locations. In this research, data was collected from mainly the QUT Samford Ecological Research Facility located north-west of Brisbane, Australia. Audio data was also used from other locations in Australia, which include the greater Brisbane area, Groote Eylandt in North Queensland, as well as St Bees Island located off Queensland coast near Mackay.

Audio sensors are deployed in the field and collected after an allotted time. Typically, the sensors that are deployed are standalone devices that record near-continuously for several days. The majority of the sensors used are *Song Meter SM2+* devices made by Wildlife Acoustics (http://www.wildlifeacoustics.com/). Custom-built sensors, with modified voice recorders are also

sometimes used. For the SM2+s, when equipped with extended batteries or solar panels, the devices can record constantly until they run out of storage space. They can record continuously for up to 10 days on a normal set of batteries and store data on multiple SD memory cards. The most common audio format the sensors record with is compressed MP3, with a sample rate of 22 050Hz, and a 128Kbps constant bit rate. A day (24 hours) of audio in this format is approximately 1.3GB in size.

Data is periodically retrieved from the sensors by field workers. A sensor is retrieved, replaced, or the memory cards and power supply are swapped out. The data is then transferred physically via mail or in person to a high-bandwidth site (usually within the same network of the destination server) and uploaded to a central sever for cataloguing and storage. Once integrated with the system, processing and analysis of the data is then possible.

Further detail about the sensors and the deployment process can be found in the publication (section 3.9).

3.2 Playback of Audio

The audio data is made available through a website for playback and analysis. Just listening to the audio data interests some ecologists; however, the size and opaqueness of raw audio data means it is of little use to ecologists attempting to infer large-scale patterns. Before ecologists can benefit from the recorded audio, the audio must first be analysed (either manually or automatically).

A website was created for ecologists and participants and has features to organise and analyse audio data. There are sections for managing *Projects* (logical collections of sites), managing *Sites* (collections of audio from a geographical location), viewing audio, searching for audio, uploading audio, discussion forums, and an area for setting up batch jobs for running automated analyses. Importantly, the annotation editor is a dedicated tool in the website that allows analysts to view audio and annotate acoustic events (Figure 9).

The annotation editor has several key components, the most basic of which is to playback collected audio data. In addition to auditory playback, the acoustic data is visualised as a spectrogram.

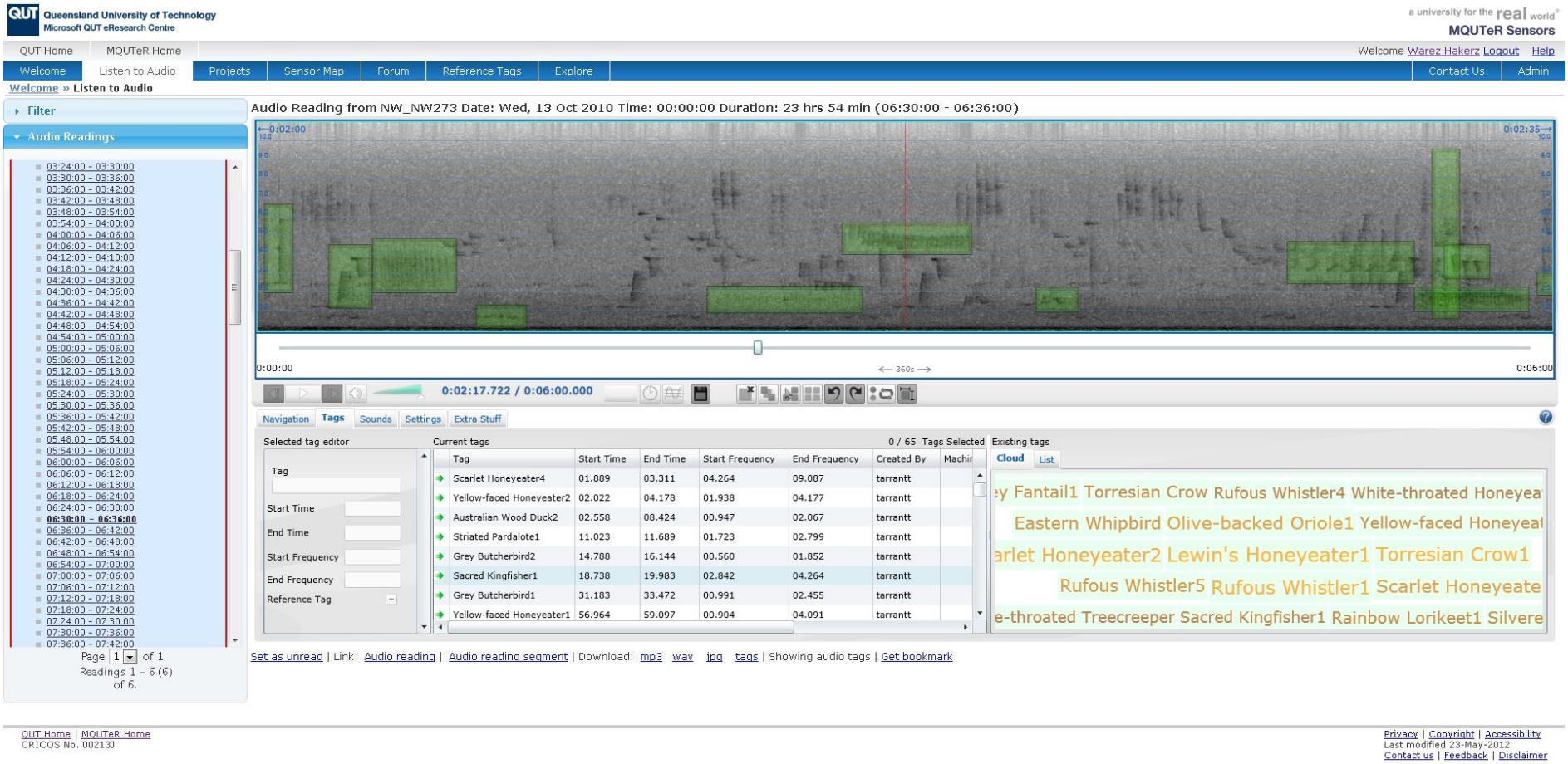


Figure 9 – A screenshot of the annotation editor

3.3 Analysis

The collection, retrieval, and to a lesser extent, the accessibility of the data are solved problems. The bulk of research remaining is related to analysing the large amounts of data collected. The aim of the analyses is to detect and classify faunal events within the recordings. There are three main forms of analysis: manual methods, automated methods, and a combination of both methods termed *semi-automated analysis*.

Processing environmental acoustic data is a *big data* problem. Audio data is collected from a multitude of sensors every day, resulting in a data collection rate that is greater than real time. Typically, sensor data is retrieved en masse – months of data become available after collection. Audio data is not as large video data but its time resolution is greater; video data is typically recorded at 60 Hz (60 FPS) whereas audio data is recorded at various sample rates, from 8 kHz to 44 kHz. Together, these fundamental criteria make large-scale processing of environmental audio data a big data problem.

3.3.1 Manual Analysis

Analysis of audio from sensors could be done manually. For example, a notepad could be used to record instances of acoustic events and audio playback applications could be used to listen to the audio. However, the human effort that would be required makes the task nearly impossible. Most standard playback software is not designed for analysis. Often no spectrogram can be displayed and there is no accurate way to measure an acoustic event's bounds (in either time or frequency domains). These problems could be addressed with the use of a more advanced program like Audacity (Audacity Team, 2013). Audacity however, is a sound manipulation tool, and not well suited to the playback of large audio files.

The hardest part of a manual process is the effort required to standardise annotation (Greenwood, 2007). Most manual processes differ on the types, frequency, and rules of annotating acoustic events. Even when these parameters are defined by the project, human analysts can still deviate from the specified behaviour. Variation in manual processes also manifests in the naming conventions used to label results. Additionally, result data is inherently distributed and needs to be collated for aggregation and distribution.

3.3.2 Automated Analysis

The best method of analysis is a fully automated approach. The ideal solution would have researchers design algorithms that take an input recording and output a list of species for all fauna that vocalised within the recording. This kind of solution does not rely on human analysis and thus reduces inconsistencies and speeds up analysis. To further increase processing speed, automated

analysis can be scaled with compute-resources – like High Performance Computing facilities / Compute Clusters / Cloud Computing. Although a fully automated approach is ideal, it is currently considered an intractable problem.

Many problems make a fully automatic solution difficult: audio data is computationally complex, soundscape and faunal vocalisation vary geographically, and existing speech-recognition techniques do not always apply. Existing techniques for human speech recognition are generally not applicable because parameters that are consistent for human speech, like a constrained lexicon, low SNR, repeatable input, and types of syllables, are not consistent for fauna.

Smart sensors that process audio data as it is recorded, possibly using the output to only record relevant data, have been investigated. Methods for the automatic detection of fauna are still under development and cannot be deployed on sensors. Simpler metrics (loudness, SNR, Zero-crossings, ACI (Pieretti et al., 2011)) can also be used. As an example, the Songmeter SM2s can be configured to record only when the input signal remains above a set amplitude. However, the variations that occur in soundscapes mean that settings for one area are generally not applicable to others. The simplest reliable solution is to record everything, so ground truthing can be repeated on source data at any time. In addition to general problems with analysing audio data, substantial time, money, expertise, and testing are required to build an automated analysis. There are two basic approaches to automatic faunal event detection: generalised recognisers and species-specific recognisers.

A generalised recogniser is a one-size fits all recogniser that could ultimately detect all target acoustic events. This all-in-one solution is currently unfeasible because it is very complex. Other researchers are actively pursuing the idea of a generalised recogniser.

The other major advantage of generalised recogniser is that ideally, it could be non-parametric. This means the algorithm does not need to be tuned for each use on a different species. Rather it could rely on only training data to adapt and operate on new sources of audio data.

This type of recogniser is extremely difficult to develop and as such, no complete solution has yet been developed. To operate effectively, a robust set of features is needed that are capable of describing every possible valid faunal acoustic event.

This method also needs a substantial amount of training data to be effective. Before operating on recordings from any new geographical area, at least some of the acoustic events must be labelled manually first (bootstrapping).

A generalised recogniser is likely to have low accuracy or precision until very well trained. This is acceptable because it can be combined with a semi-automated approach while it is still being trained.

Research is also devoted to generalised recognisers that cluster similar acoustic events together. This unsupervised method of analysis would need little training data but would require clusters to be labelled – perhaps again in a semi-automated fashion.

Alternately, species-specific recognisers are designed to detect one species or one type of acoustic event at a time. These algorithms are designed per species; they are modular, and sometimes simple. This approach of constructing a recogniser per species is inefficient. Because they are built separately, using the appropriate methods for each type of species, their implementations are often different. Different implementations make it difficult to apply the whole collection of recognisers in one pass of the audio data.

Species-specific algorithms can often be fine-tuned since they are usually parametric – the variables that control the algorithm can be precisely adjusted to accommodate for small variances in the target acoustic pattern. Because of fine-tuning, each recogniser can potentially be very effective. The downside to this approach is that adjusting parameters can require an expert.

Because each recogniser needs to be designed to use features that best represent the target species, only rarely is reuse of an algorithm for another type of vocalisation possible. Additionally, anecdotal evidence suggests, in an academic environment, the time taken for each recogniser to be implemented and fully tested can take between one to three months (M. W. Towsey & Planitz, 2010). An expert in recogniser design must be employed to develop these recognisers; this is an expensive scenario.

When the QUT Ecoacoustics research group was first formed, there was no training data available. Thus, the development of a generalised recogniser was not possible. Instead, the research group has successfully developed several species-specific recognisers: recognisers for the koala, kiwi, canetoad, ground parrot, lewin's rail, gastric brooding frog, and others.

3.3.3 Semi-Automated Analysis

An alternative to both the manual and automatic approaches is a combination of both. Computers are excellent at processing large amounts of data in an objective fashion. However, it is difficult to create and train algorithms for computers that are capable of detecting and discerning the complex patterns found in audio data and associated spectrograms.

In contrast, a human analyst is an excellent analyser of certain types of data. Humans are exceptional at discerning differences or similarities between two patterns, especially in audio or image data (see section 2.1.5). Humans can adapt to different problems in a creative and dynamic manner; analysts can account for noise, overlapping signals, and false similarities with relative ease. However, human participants still make mistakes and their ability to memorise data for classification tasks is limited when compared to a database. More importantly, human analysts tire, get bored, or lose motivation when given too much work.

A semi-automated approach to the annotation of faunal acoustic events in audio recordings is the complementary combination of the strengths of the human and computation approaches. A semi-automated approach is any methodology where technology is used to assist the analyser, with a full range of levels of involvement for both humans and technology. Ideally, a semi-automated approach that is beneficial to both parties would strike a balance between human effort and machine effort.

The beginning of a semi-automated system for processing data was built into the website used by this research. As shown in section 1.1, the QUT Ecoacoustics research group provides online tools for volunteers to analyse the recordings collected from the sensors (Wimmer, Towsey, Planitz, Williamson, & Roe, 2012).

The annotation tool has an interactive drawing surface that allows a participant to marquee an acoustic event of interest and associate a descriptive, textual, tag with it. An *annotation* is defined as the combination of an acoustic event with one or more tags. To marquee an acoustic event, the participant draws a rectangle around the event on the spectrogram to specify the time and frequency bounds of the event. Each bounding rectangle should be labelled with a tag representing the common name of the species that was responsible for the vocalisation. The tag data, the marquee's bounds, the time of day, the location, and the species that called, are the core pieces of data generated by analysis.

3.4 The Faunal Acoustic Event Annotation Process

Annotations are an important part of the data output by the analysis of audio data. They can be generated by human analysts or through automated methods. This section provides a short overview on the steps involved for annotating, with perspectives from human and machine analysers. Generally, to annotate an acoustic event, three steps are required (refer to Figure 10, page 59 for a labelled diagram):

1. Detection: Find an acoustic event – an Event Of Interest (EOI)
2. Segmentation: Specify the bounds of the event, defining what signal is part of the event

3. Classification: Classify the event by applying a discriminating label to it

The first step, finding an acoustic event, for a human, consists simply of searching through acoustic data listening for significantly interesting events. When a spectrogram is shown as well the task involves scanning through an image looking for spectral segments that stand out. With modest practice, humans learn to scan through spectrograms intelligently, avoiding uninteresting spectral blobs. Searching for acoustic events is a simple but menial task; with time, analysts can become bored. Boredom affects data quality and analyst retention (see section 2.1.5).

Algorithmically, event detection is a moderately hard task. Some research simply skips EOI detection, instead focusing on the more interesting classification problem (Acevedo et al., 2009). Common methods for detecting EOIs include looking for 'peaks' of acoustic energy in spectrograms that meet certain requirements (Bardeli, 2009; Dong, Towsey, Jinglan, Banks, & Roe, 2013; M. W. Towsey & Planitz, 2010), or looking for significantly different sections in the audio data (Duan et al., 2011). Although machines perform this task quickly, they cannot automatically determine what is truly interesting. To compensate, they are often tuned to report an overabundance of EOIs resulting in many false positives.

The second step of the annotation process, segmentation, involves bounding the event. In this stage, a human participant is required to isolate the region of the spectrogram that bounds the acoustic energy from the source of the EOI – typically one faunal vocalisation. Participants are instructed to annotate entire vocalisations rather than just the component syllables of a vocalisation. Audio playback helps participants classify ambiguous events and isolate event sources (both the target source as well as other interfering acoustic sources).

Again, machines find this task challenging. Outlining the bounds of an event is not too hard: most algorithms have some form of templating, sequence matching (HMMs or Timed automata), or make use of some kind of spidering algorithm (Duan et al., 2013). The real difficulty is isolating sound sources. Determining the source of overlapping audio components is a difficult task for current algorithms.

The last step for annotation is classification of the event. Both humans and machines do this by encoding the event with a distinguishing label. In this research, tags are used. Machines will either classify a single type of event (binary classification) or classify many different types of event (multiclass classification). Both approaches can make use of training data. Regardless, machines are quick, precise, and efficient when compared to a human classifier. Conversely, the only comparisons algorithms make when classifying are defined by the attributes (features) that they are programmed

to check – unlike a human classifier, who is capable of using creative, abstract, and qualitative features.

Overall, however, humans find the classification stage difficult. The two stages a human employs to classify an event are pattern recognition and then labelling. The pattern recognition task involves a human matching the data either via memory or by example from a reference-source. If matched in memory, then the participant has encountered the pattern of data before and can usually label it (implying some kind of skill or training). If matched to an example, the label from the example is used. If a participant cannot match the pattern at all, then classification cannot be completed by that participant.

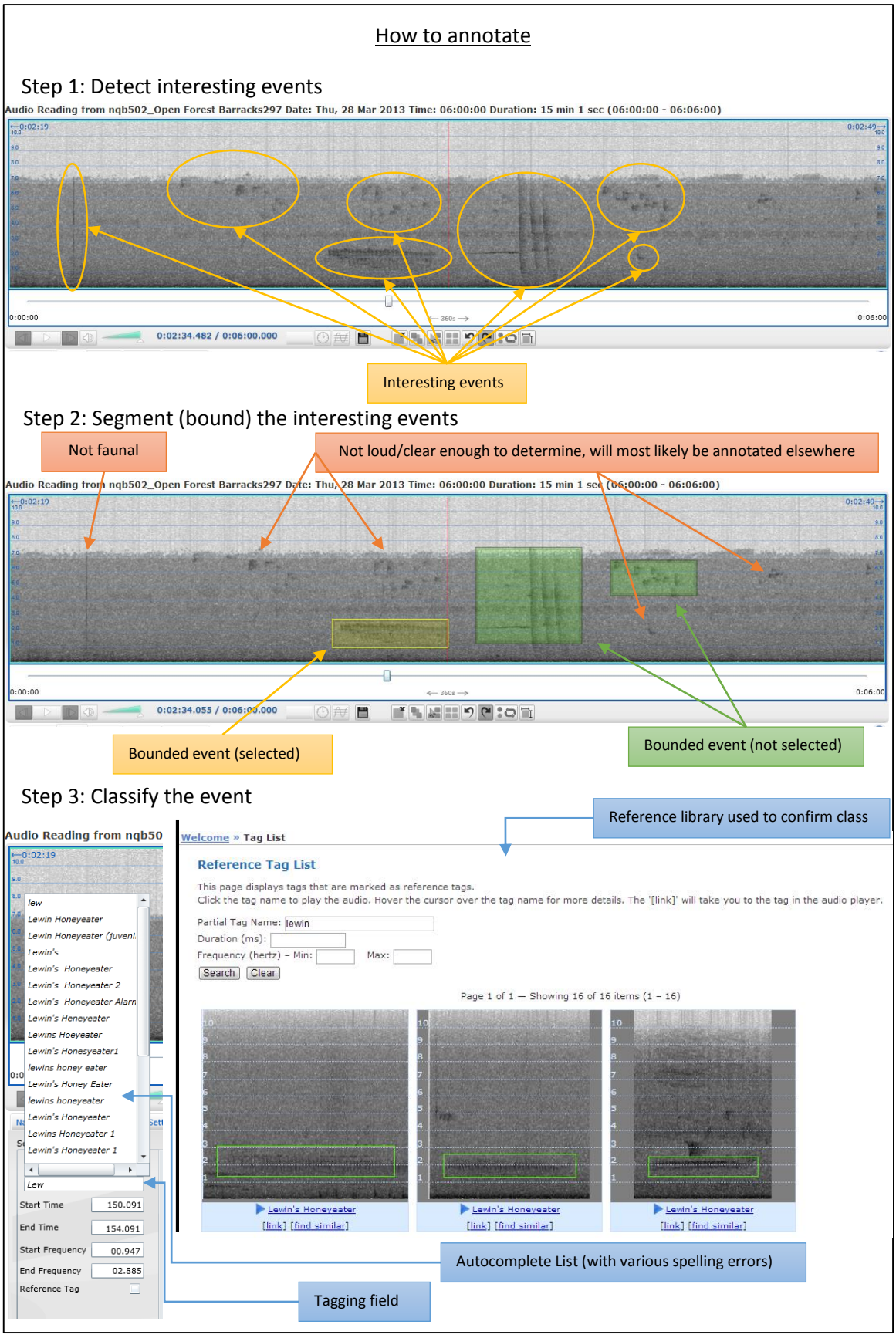


Figure 10 – A diagram depicting how faunal annotations are created on the current QUT Bioacoustics website

3.5 Data Architecture

This section describes the scale, structure, and variety of the datasets available to this thesis. The QUT Ecoacoustics website was designed to store large amounts of audio data collected from sensors. Several basic data entities hold the vast majority of the data stored. A simplified diagram is shown in Figure 11. It should be noted that the basic entity layouts listed in this section are remarkably similar (through coincidence) to the Pumilio bioacoustics project (Villanueva-Rivera & Pijanowski, 2012).

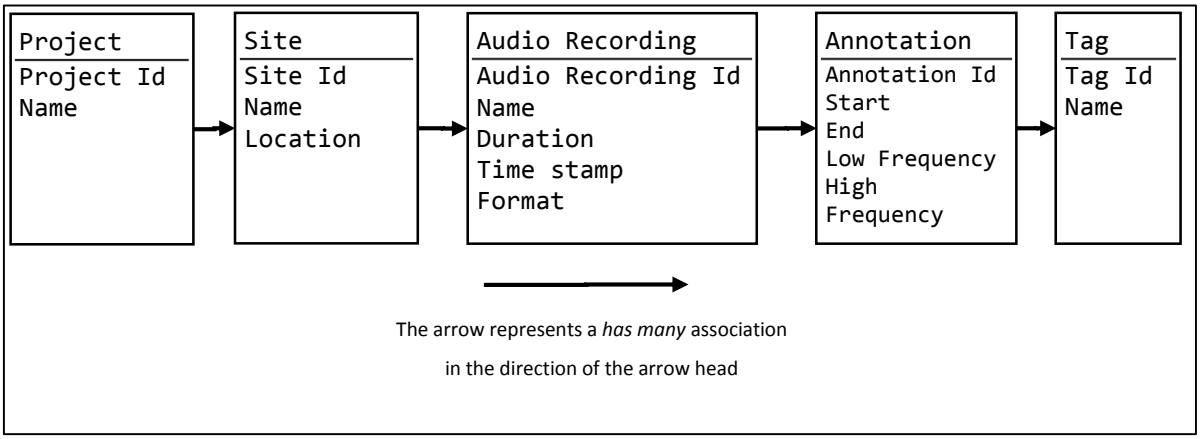


Figure 11 – A simplification (in UML notation) of the important entities in the website

`Projects` are logical containers for grouping associated data together. As an example, the 'SERF Acoustic Study' `Project` contains most of the data used in this thesis. `Projects` have many `Sites`. As an indication of scale, there are currently 35 `Projects` in the QUT Ecoacoustics Research group's database[1].

`Sites` are named after their ecological namesake. A site in ecology is a specific location where a study has occurred. An example for bioacoustics is a manual field survey for a limited transact, like the ATLAS Record Form (see section 2.3.2). For the website, a `Site` is a specific location where a sensor has been deployed. `Sites` have many `Audio Recordings`. There are 226 `Sites` in the database[1].

`Audio Recordings` are the basic unit of data collected by the website. `Audio Recordings` must be associated with a `Site` (and thus they have a location as well as belonging to a `Project`). The recording start datestamp as well as cached meta-data (like audio format, sample rate, and

[1] As of the 1st of September 2013

duration) are stored in the database. There are 238 653 `Audio Recordings` in the database with a total duration of 3.56 years of continuous recordings[1].

`Audio Recordings` have many `Annotations` – `Annotations` are the principle result of analysis of the acoustic data. The `Annotation` entities that are stored have a '`Human Made?`' attribute that indicates whether a human or an automated analysis generated the `Annotation`. `Annotations` have many `Tags`. There were 120 736 `Annotations` in the database, created between the 19th of October 2007 and the 24th of August 2013[1]. These annotations have only been applied to 4428 `Audio Recordings` This means that only 1.9% of the entire `Audio Recording` collection has received any analysis at all. Even then, the distribution is extremely skewed – 95% of all `Annotations` occur on only 941 (≈0.4%) of all `Audio Recordings`. This distribution is visualised in Figure 12.

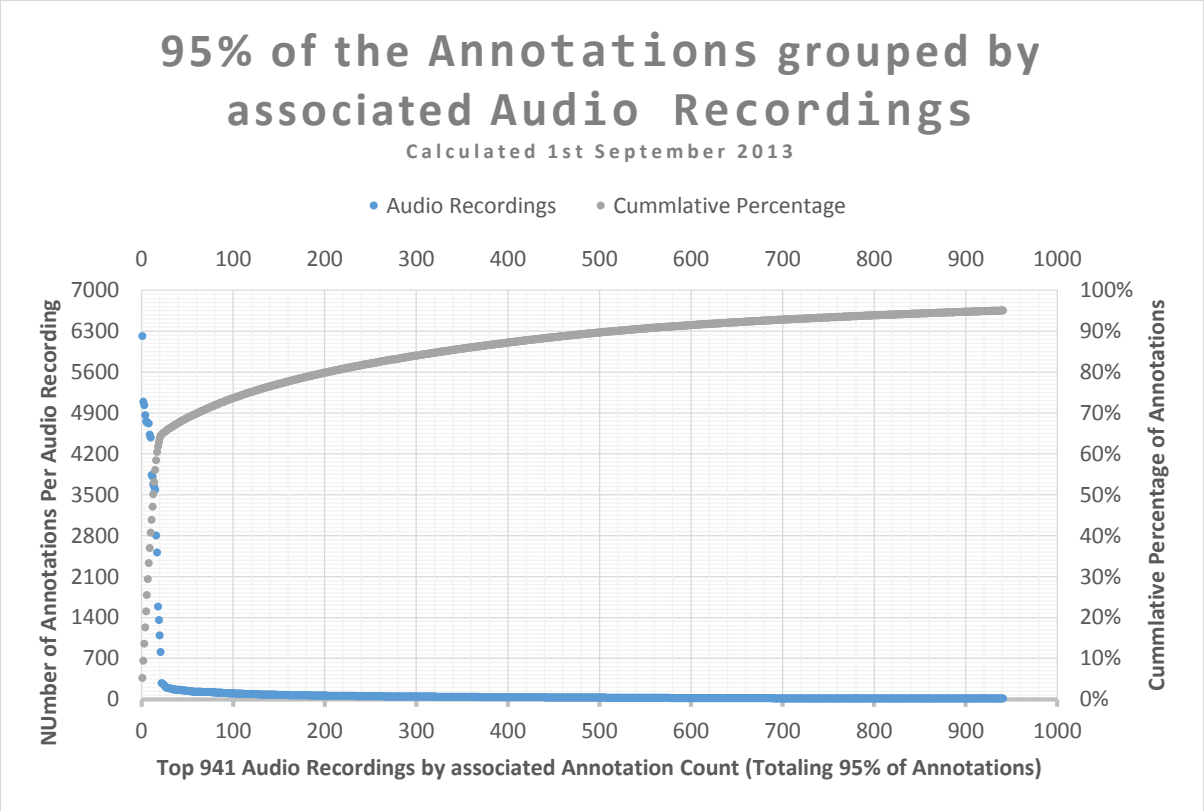


Figure 12 – A chart of the distribution of `Annotations`, ordered by `Audio Recording` association density

It is expected, if all `Audio Recordings` had received full analysis, that there would be between 6.4 million and 918.6 million `Annotations` in the database[2]. The statistics highlight the collection/analysis imbalance present in the system (previously discussed in section 1.1).

`Tags` are simply short descriptive textual labels that can be applied to an entity. Unlike the previous exponential increases in scale seen by the other entities, there are (relatively) few `Tags` in the database. There are 1398 `Tags` in the database[3]. This statistic includes hundreds of duplicates, misspellings, and erroneous tags. These errors are dealt with in section Chapter 7 (*Tag Cleaning and Linking*).

The entities described above are simplified versions of the true entities stored, which have many more tracked attributes. In addition, the database stores other types of entities, most of which are linked to the principle entities previously described.

3.6 Annotation editor

Each of the principal entities above are managed by basic web interfaces. Most of these entities only need basic user interfaces for management, known commonly as CRUD operations (create, read, update, and delete). Screenshots of these interfaces are included in *Appendix F – Annotation Software Platform Screenshots*.

However, the entity that needs the most specialised user interface is the annotation entity. The annotation data the system produces (a list of species, call-types, times, durations, and frequency bounds, at a location) is the core output data for the research group; it is used by ecologists.

At its core, the editor is an audio playback tool. Based off the functionality offered in Audacity (Audacity Team, 2013) and seen in other software packages (Raven, Song Scope, and most audio manipulation software) the playback of audio is accompanied by a spectrographic visualisation.

In a semi-automated annotation model, there must be a way for a user to create and manipulate annotations. Thus, in addition to playback functionality, the annotation editor allows for the viewing, creating, and editing of annotations – either human or machine generated.

[2] The minimum estimate was calculated as the average number of Annotations per Audio Recording (≈27) multiplied by the number of Audio Recordings present. The maximum estimate was calculated as the average number of Annotations per Audio Recording, for the top 20 most annotated Audio Recordings (≈3850) multiplies by the number of Audio Recordings present.

[3] As of the 1st of September 2013

The original annotation editor was Microsoft Silverlight control, embedded in a simple web page. When created in 2009 there was still heavy focus applied to completely automated methods of analysis.

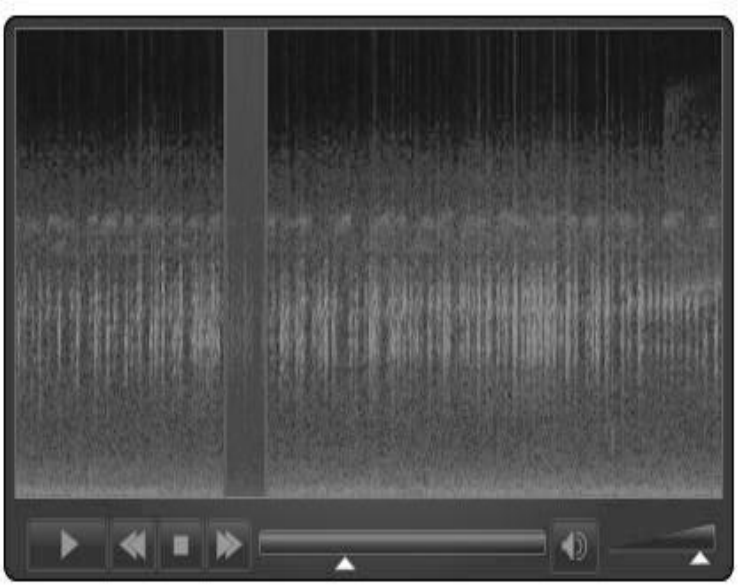


Figure 13 – A screenshot of the original annotation editor (Mason et al., 2008)

However, in response to the increasing need for data, a more sophisticated version of the 'player' was created, which could perform as an interface suitable for prolonged human input. The improvements, seen in (page 64), were based off feedback from the human analysts using the system.

Visible changes include: frequency lines for the spectrogram; adjustable window width, with time bound labels; frequency bounded annotations (the original version only bounded annotations by time); the capability for multiple annotations to be associated to a recording; a built-in differential saving mechanism; advanced annotation editing abilities, including functionality for: deleting annotations, sending an annotation to the back/font of the z-stack, selecting/deselecting all/single/multiple annotations; multi-annotation editing; detail and list views for annotation editing; autocomplete and tag cloud assistance; an initial implementation of the suggestion tool (see Chapter 5); a 100% action coverage undo/redo stack; configurable personal settings; a built-in reference library of exemplar annotations; the ability to show immutable annotations; and data exports links.

Non-visible changes include support for the playback and editing of long recordings (from originally 2 minutes, to support for recordings up to 48 hours long); segmentation and preloading techniques; a dedicated caching mechanism; and the ability to read audio data from local storage (as opposed to

always needing to download the audio from the internet). The last two points were particularly important for analysts that had both limited internet bandwidth and quotas.

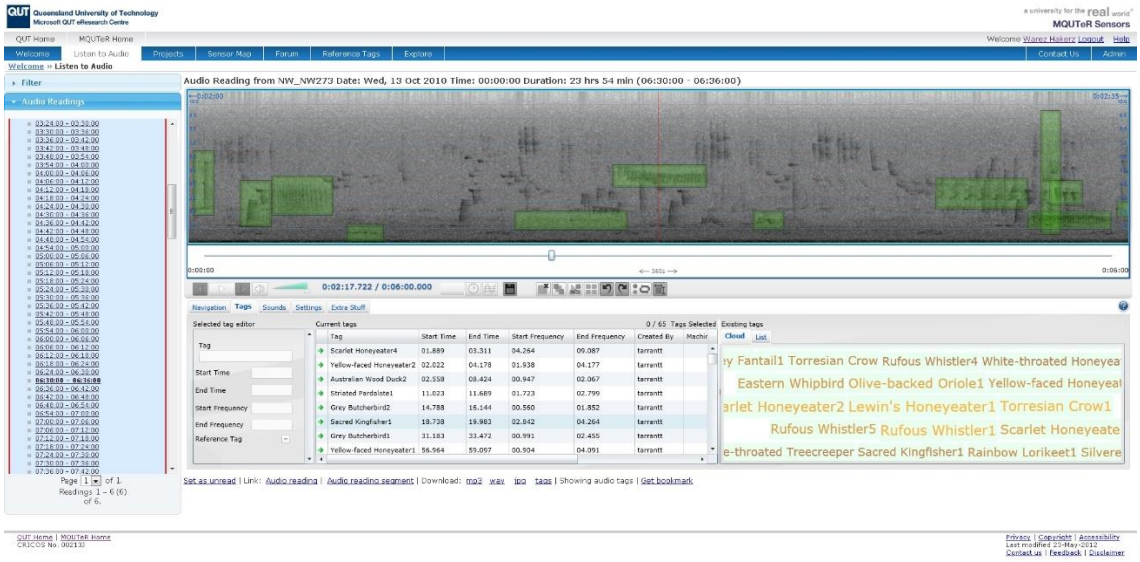


Figure 14 – A screenshot of the improved annotation editor

3.7 Open Source Efforts

Starting in October 2012, the software platform was redesigned and re-implemented. The new software package is open source and available at https://github.com/orgs/QutBioacoustics[4]. The decision to re-implement the system was made for several reasons:

An open source model was chosen based on feedback from bioacoustics workshops. By some metrics, the advanced development of the QUT Bioacoustics software platform makes it a research leader. Other bioacoustic software platforms (particularly Pumilio and ARBIMON) are have similar features sets when compared to the QUT software package. Other researchers are seeking pre-made alternatives rather than building their own (expensive) software packages. Open sourcing the new code base allows contributions from the larger community.

Years of development of the previous software platform resulted in unnecessary complexity in the architecture of the system. The complexity accrued as new features were added experimentally. Some of these features were unnecessary and later discarded, resulting in confusing stubs and

[4] Every effort will be made to ensure this link is permanent. However, it may eventually be taken down from Github's hosted repository. In which case, the author of the thesis has a copy of the relevant source code and a copy can be obtained by correspondence.

unnecessary abstraction. Other new features were kept but could have been designed better. Some unnecessary features were never removed because of their very complex coupling to other features.

Since its inception, the research group has learnt much about designing bioacoustic software. Implementing the website, in a simpler, more streamlined manner, unrestrained by backwards compatibility, was ideal. The resulting simpler architecture makes the software more accessible for open source contributions.

The previous website required Silverlight and Flash to be installed on a client's machine. The new platform was built using HTML5 technologies for the client interface. Any reasonably new device with browsing capabilities can run the annotation editor and there are no dependencies on browser plugins (like Flash or Silverlight). The new website supports modern browsers, including Google Chrome, Mozilla Firefox, and Internet Explorer (version 10+).

3.8 Conference Paper – Practical Analysis of Big Acoustic Sensor Data for Environmental Monitoring

This paper published the details of the applied research used in the QUT Ecoacoustics research group. Much of the information within is relevant to this thesis. In particular, it describes the methodology used to gather and process acoustic sensors data.

Truskinger, A., Cottman-Fields, M., Eichinski, P., Towsey, M., & Roe, P. (2014). *Practical Analysis of Big Acoustic Sensor Data for Environmental Monitoring*. Paper presented at the 2014 IEEE Fourth International Conference on Big Data and Cloud Computing (BdCloud), Sydney, Australia. http://dx.doi.org/10.1109/BDCloud.2014.29

This conference paper has been peer reviewed and published.

3.9 Statement of Contribution

RESEARCH STUDENTS CENTRE
Examination Enquiries: 07 3138 1839
Email: research.examination@qut.edu.au

Statement of Contribution of Co-Authors for Thesis by Published Paper

The authors listed below have certified* that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter:

Publication title and date of publication or status:

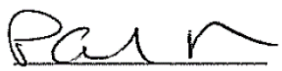
Practical Analysis of Big Acoustic Sensor Data for Environmental Monitoring

Published, 2014

Contributor	Statement of contribution*
Anthony Truskinger	Co-wrote the manuscript, co-created the software, methodologies, and frameworks presented in the paper.
Signature 	
Date 21/05/2015	
Mark Cottman-Fields*	Co-wrote the manuscript, co-created the software, methodologies, and frameworks presented in the paper.
Philip Eichinski*	Co-wrote the manuscript, assisted with the software, methodologies, and frameworks presented in the paper.
Michael Towsey*	Assisted with the manuscript, assisted with the software, methodologies, and frameworks presented in the paper.
Paul Roe*	Supervisor – Oversaw and contributed to the entire paper

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Paul Roe		20/5/15
Name	Signature	Date

Practical Analysis of Big Acoustic Sensor Data for Environmental Monitoring

Anthony Truskinger, Mark Cottman-Fields, Philip Eichinski, Michael Towsey, Paul Roe
QUT Ecoacoustics Research Group
School of Electrical Engineering and Computer Science
Queensland University of Technology
Brisbane, Australia
{anthony.truskinger, m.cottman-fields, phil.eichinski}@student.qut.edu.au, {m.towsey, p.roe}@qut.edu.au

***Abstract*—Monitoring the environment with acoustic sensors is an effective method for understanding changes in ecosystems. Through extensive monitoring, large-scale, ecologically relevant, datasets can be produced that can inform environmental policy. The collection of acoustic sensor data is a solved problem; the current challenge is the management and analysis of raw audio data to produce useful datasets for ecologists.**

This paper presents the applied research we use to analyze big acoustic datasets. Its core contribution is the presentation of practical large-scale acoustic data analysis methodologies. We describe details of the data workflows we use to provide both citizen scientists and researchers practical access to large volumes of ecoacoustic data. Finally, we propose a work in progress large-scale architecture for analysis driven by a hybrid *cloud-and-local* production-grade website.

Keywords—acoustic sensing; bioacoustics; data analysis; scalable analysis; cloud infrastructure; ecoacoustics

I. INTRODUCTION

Sensors are an effective tool for the large scale monitoring of the environment. Acoustic sensors are regularly used to monitor vocalizing fauna with the intent of assessing biodiversity [1, 2]. Acoustic sensor data can also address ecological questions relating to the vocalizing patterns of fauna, the presence or absence of species, and species abundance. The volume of data generated by sensors requires large compute resources for analysis. This paper elucidates the practical analysis methodologies that will allow for a hybrid *cloud-and-local* compute architecture required by our ecoacoustics project.

Traditional methods of surveying ecosystems are manual and require field workers to visit the site of study. While the results of manual surveys remain valuable, sensors have several advantages: they record data constantly, cost relatively little, are minimally invasive, and create a permanent, objective record of a site. Deploying sensors over large spatiotemporal scales allows scientists to collect massive amounts of data.

Advances in sensor technology, specifically in storage capacity, in the last 10 years, have provided the hardware for practical large-scale collection of data. The Wildlife Acoustics' SM2+ [3] is a commonly used acoustic sensor [4-7] that can be deployed with four high density SDHC cards and an external power supply. A solar-powered SM2+ sensor can record audio for over a year (128kbps MP3, 1024GB storage). With reliable sensors and high-density storage, collecting data is no longer considered problematic. Instead, ecoacoustics research now concentrates on the questions of managing and analyzing ecoacoustic data; the latter of which is a more complex and varied problem [8].

Automated methods of analyzing acoustic data are preferred; however, currently there exists no single, generalized, automated solution for identifying all vocalizing fauna within sensor audio recordings. There are two broad reasons for this intractability. First, automated identification of species is difficult due to the variability that faunal vocalizations exhibit, the low *signal to noise ratios* (SNR) endemic to acoustic sensors, and the acoustic competition between species that adds further complexity to the data [1]. Second, practical methods for analyzing, visualizing, and understanding acoustic sensor data are still not well developed. Raw audio data is opaque and hard to reason about without analysis [9, 10].

Analysis and management of ecoacoustics is a big data problem and our research to solve this problem has produced software artifacts such as the Ecosounds Acoustic Workbench (pictured in Fig 1). Employing the 5Vs of big data [11-13] as metrics, the QUT Ecoacoustics Research Group collects data that has:

- Volume: Currently, 24TB of acoustic sensor data has been collected. Of that, 15TB has been ingested into the Bioacoustic Workbench – a production website – where audio can be accessed (navigated, played, and shown as spectrograms) on demand.
- Velocity: The research group has access to 50 sensors; there is a potential data velocity of 355GB/day (Stereo WAVE, 22050Hz, 16-bit samples).
- Variety: While sensors produce data in consistent formats, the content can vary wildly over small geographical distances. Techniques applicable to one region often do not work in others. Additionally, various methods of analysis produce many types of data, including visualizations, indices, events, points of interest, spectra, metadata, annotations, or tags. Processes that involve people performing analysis can introduce further variety.

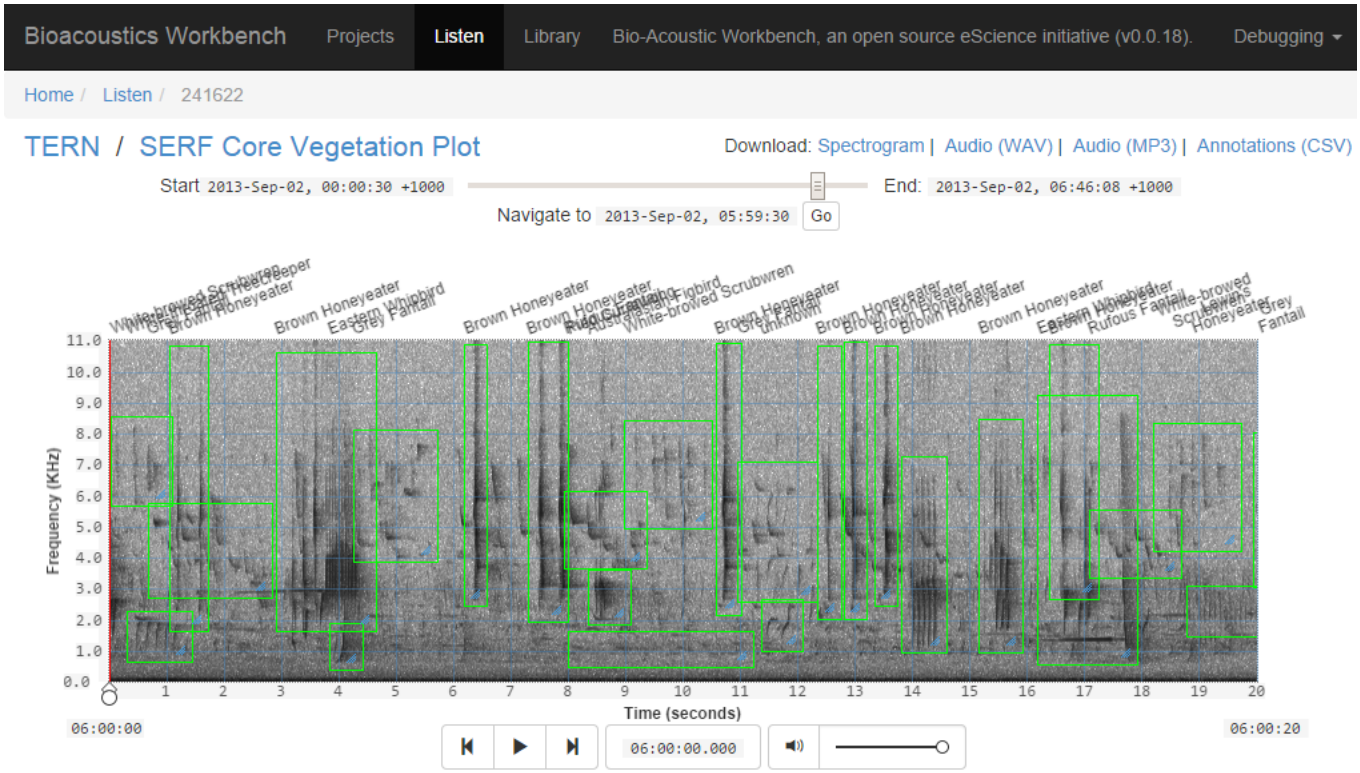


Fig. 1. A screenshot of the Ecosounds Bioacoustic Workbench's annotation interface

- Veracity: The raw data produced by sensors are an objective record of activity – this is an inherent advantage of using sensors over manual studies. However, human-driven analysis or the verification of automated analysis creates potential sources of data uncertainty.
- Value: The results from collecting and analyzing acoustic sensor data can produce valuable ecological data for input into the formation of environmental policies.

This paper presents software, methodologies, and supporting architecture for analyzing large sets of acoustic sensor data. Scientists within our research group and external collaborators have made use of the processes and software described by this paper. Our contribution is to publish our applied large-scale analysis research, details of our migration to cloud based architecture, and our open source software to aid other researchers in the field. Related work is presented, followed by an overview of the acoustic sensor data workflow. Then, a detailed report on methodologies is presented. Finally, a work in progress section details plans for scaling up the analysis architecture.

II. Related Work

There are a growing number of data intensive projects with varying research foci. Within data-intensive science, there are recognized differences in dataset sizes, computational needs, and collaboration standards. Our work is firmly in the middle of Jim Gray's long tail of science [11]. Large-scale ecoacoustics requires reasonably complex technology, as well as computer scientists and IT experts to manage and process data [14]. The volume of data being processed necessitates an evolution beyond spreadsheets, flat files, and hand-curated data – the methods of independent scientists.

While most audio datasets are not equivalent in size to genome or astronomy data (typically in the petabyte range) [15], terabytes of audio still pose a significant challenge. Volume on disk does not necessarily equate to complexity in processing. Acoustic data is opaque and by definition always represents data over time. This makes it difficult to summarize, visualize, or even manually preview individual files [10]. Effectively characterizing local areas as well as large amounts of data, obtained across large spatiotemporal periods, is challenging. Analysis of acoustic data using indices and broad methods of comparison and differentiation have been used to successfully obtain an overview for comparing acoustically similar areas [4].

Recordings of fauna vocalizing are commonplace. However, there is an important distinction to be drawn between targeted recordings and untargeted recordings. Targeted recordings, also known as trophy recordings, are usually short, contain just one call, have a high SNR, and are usually captured with specialized equipment. These recordings have a relatively low cost in terms of data volume and analysis complexity. Untargeted or general environment recordings, like those produced by acoustic sensors, are typically very long (hours to days per recording), have many vocalizing fauna, low SNRs, and can capture overwhelming amounts of irrelevant signal and

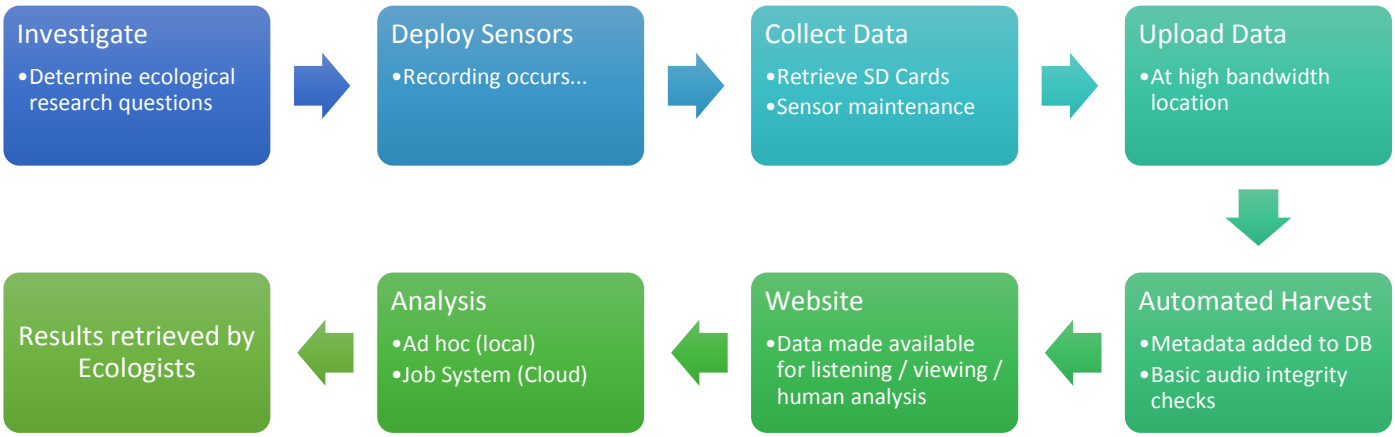


Fig. 2. The QUT Ecoacoustics Research Group's process for collecting data from sensors

background noise. These recordings have a high cost in terms of volume of data and analysis.

The Xeno Canto website is a collection of faunal vocalizations in targeted recordings. The majority of recordings are short, with a high SNR. The site has similar goals to our project – increasing the data available on the environment and biodiversity – with a vastly different approach. The short recordings lend themselves to manual listening and analysis. It is possible to discuss an entire recording and often be sure of which sound source is the 'target' of the recording. Xeno canto currently has approximately 500GB of audio recordings [16]. Sensors however, generate very large, untargeted, recordings – it is not feasible to discuss or analyze that data with Xeno Canto methods.

There are a number of commercial programs that can be used to analyze acoustic sensor data to detect vocalizations of interest. SongScope and Raven are two programs that can achieve reasonable accuracy in smaller audio datasets with supervised training [17]. Unfortunately, neither of these programs are designed to scale to very large datasets.

Pumilio is a successful open source ecoacoustics web application [18]. It has multiple deployments actively used by different research groups, allows for uploading, listening, and analyzing audio. The project has focused on easy deployment and use. Pumilio is designed to run on a single machine – possibly in the cloud – it is not clear how the project will deal with significant scale.

III. Methods – Data Collection

This section details the methods employed to gather acoustic sensor data by our group. This process is depicted by Fig 2.

Initially, ecological research questions are provided by collaborating ecologists, community environment groups, businesses concerned about their impact on the environment, or government initiatives. The research questions utilize acoustic information from sensors, sometimes indirectly, to form conclusions.

Sensors are deployed into the field in different configurations. Typically, recorders are placed at ecotones (sites that are a transition between two biomes) to maximize the variety of species detected. Sensors can also be deployed to target specific species or in patterns (like grids). Factors that affect sensor performance include territory size of targeted fauna, vocalization amplitude & frequency of target fauna, vegetation type, terrain, and environmental noise sources.

SM2+ sensors (Fig 3.) are the most commonly used; they can potentially record audio unattended for over a year. However, we typically employ one of two patterns: weeklong or four-month long cycles (deployed for up to 3 years). These shorter cycle times allow data to be incrementally gathered. When the data is gathered, health checks and maintenance are also conducted. Weeklong cycles require four D-cell batteries, whereas the four-month cycles (≈125 days) are deployed with a solar panel and a deep-cycle battery. Both types of deployment record data in a stereo WAVE format (PCM, 22050Hz, 16-bit samples). The SM2s have two microphone inputs – utilizing

Fig. 3. A deployed SM2+ Sensor

both microphones creates redundancy in the event of a single microphone failure.

At the end of a cycle, a field worker will inspect a deployed sensor. If it is the end of the deployment, the sensor is retrieved. If a deployment has not concluded, the SD cards are swapped out. Regardless, the cards are physically returned to a high bandwidth location (typically within a university's network) and the data is uploaded to a working area. When metadata files are added to each directory, an automated harvester detects the changes and schedules harvest jobs for each waiting audio file. Files are converted from WAC if necessary to WAVE – other file formats do not require pre-ingestion conversion. The file type *WAVE* is used for uncompressed files and *WAC* is Wildlife Acoustic's proprietary lossless audio compression format.

Required analyses, either automatic or semi-automatic, are conducted before the results are sent off to ecologists. Semi-automated analysis is done by annotating faunal vocalizations [1].

IV. Methods – Analysis Development and Execution

We are an eScience research group. Our goal is to provide computer science support to traditional scientists. Nevertheless, even within our group we hire/require specialist IT professionals in addition to research staff. We propose that the concept of eScience requires graduated levels of professional IT support for data intensive science; some groups may only need small amounts of professional support, others may need small workforces (e.g. the Square Kilometer Array project [19]).

A. Developer / Researcher Tension

There is tension between the goals of researchers and software developers. As an eScience group, we regularly work with research and professional staff. One core goal of the research group is to incorporate analysis algorithms and processes into the public production website. This requires a reasonable understanding of the source code and a fixed feature base. Contrast this with the typical methodology for research work: researchers are never done improving their results and are constantly tweaking source code. Without freezing core features and APIs, it is difficult to maintain working production code [20, 21].

We have approached this problem in two main ways: Refactoring checkpoints (freeze feature sets that researchers have stopped working on) and ad hoc analysis systems.

The first concept, freezing features is a common practice in software development. In order to ship a product, new features will not be allowed, existing features will have their APIs frozen, and the only continuing work will be maintenance. A full feature freeze is not compatible with a researcher's set of priorities.

As an alternative, every few months, time is allocated for refactoring analysis code. Features and APIs that have not changed recently are marked as 'production stable' and can then be depended on. Features that are part of active research are tracked but not altered. The result is a limited but progressive set of restrictions to the researchers. This semi-regular iteration cycle works well because all parties involved know and have input into the process. The result is a naturally forming framework that adapts as analysis algorithms are developed, tested, and become stable.

The second concept we have employed is ad-hoc analysis systems, which have proven very useful. We have reserved dedicated compute resources and have some generalized scripts for running ad hoc analyses. These scripts require an IT professional to run but do not require production-level feature freeze.

B. Compute Resources

We have three basic compute resources available:

- QUT's High Performance Computing (HPC) support
- a dedicated big data processing lab (BigData) containing powerful standalone computers designed for researcher experimentation
- Queensland Cyber Infrastructure Foundation (QCIF) and the National eResearch Collaboration Tools and Resources (NeCTAR) provide access to cloud storage and cloud compute resources for data-driven collaborative research.

Our research group currently has two storage options with 100TB in total through the QUT HPC and QCIF. The two storage locations have mirrors of all audio data. In addition to serving as backups, it allows either QCIF Cloud or QUT HPC compute resources to run analysis with on-site data access. We would prefer solutions that remove the need to transfer data [22]; however we currently remain dependent on high-speed links between data stores.

The transfer of data that involves disk or network I/O generally has the largest impact on analysis efficiency. The main method we employ to reduce the required data transfer is command-line audio manipulation tools that can seek smartly through audio files. For example, *mp3splt* can segment MP3 format files without needing to read the entire file. Early in the research group's development of analyses, the amount of data stored in RAM caused paging and extreme contention for resources. This limitation has been bypassed through audio file segmenting.

The next most limiting factor is the number of processing cores. A 'big data' lab provided by the university contains twelve machines (dual Intel Xeon E5-2665, 32 virtual cores, 256GB DDR3 RAM, 3TB SCSI Raid, dual 1Gb Ethernet) designed to address the needs of researchers working with data that is impractical to process on their personal computers. Their prime benefit to our research group is unrestricted access and resulting flexibility. We also make use of their high throughput and large amount of RAM. In particular, RAM disks for storing the cache of intermediate audio files cut for each segment of analysis are very useful.

Similar to compute-cloud-based VMs, the BigData machines are used to run experimental, ad hoc analyses on demand. Although QUT's HPC facilities provide magnitudes more processing power, they also require additional structure and enforce extensive restrictions that often conflict with the development of an in-progress algorithm or research

exploration. The BigData machines have been used to produce over 8TB of analysis results. When an analysis becomes stable and the scale of the data that is produced is increased, QUT's HPC compute resources are preferable.

C. Analysis

We have several forms of automated analysis categorized into two large groups: event detection and acoustic index generation. Event detectors produce time and frequency bounding boxes around spectral components of interest in an audio signal. Event detectors have been developed for a number of species: koalas (male), frogs, cane toads, cicadas, ground parrots, crows, kiwis, Lewin's rails, as well as generalized event detectors like Acoustic Event Detection (AED) and Ridge Detection [1, 23]. Acoustic indices, in contrast to detecting faunal events in audio streams directly, instead calculate summary statistics from the audio stream to provide large-scale insight into normally opaque audio.

Almost all analyses we produce are programmed in either C# or F#. C# is an unusual choice for research programming. However, contrary to the stigma of being too expensive, significant amounts of the C# and .NET toolchain have become free in recent years. C# has reasonable speed profiles, good tooling support, includes static analysis, and has automated garbage collection. It has a C-like syntax which is beneficial to researchers with a background in C or C++. The advent of multi-operating system support through the Mono project (http://www.mono-project.com/) has allowed our analyses to run on Unix/Linux operating systems. Where the performance of C# does not match that of native libraries (e.g. those written in C or C++), for critical operations our codebase will call native versions of the required functionality. For example, Fast Furrier Transforms (FFTs) are calculated by a native library for all of our analyses. Optimizations are implemented only when necessary as indicated by profiling.

The R language for statistical computing is used for the initial exploration of datasets. We have run large-scale data analysis in R; however, after the initial research stage has ended, often the research artifact transcoded to C# for ease of maintenance and extension by our researchers. Intensive or complex audio work is delegated to specialized programs, such as *SoX*, *FFmpeg*, *mp3splt*, and *shntool*. These programs are cross platform, provide a scriptable command line interface, and operate on files. We have wrapped these tools in two dedicated APIs – one for .NET and one for Ruby programs. Our Ruby audio-tools wrapper is open source (https://github.com/QutBioacoustics/baw-audio-tools).

Reproducibility of experiments and provenance of data are encoded in the tools and processes we use. Source audio data is considered immutable, with provenance maintained through log files and database metadata. Each compilation of the analysis programs includes the Git (a distributed source control application) commit hash. This provides a direct link from results and log files back to the source code that was used. All configuration files, output from analysis, and log files for each analysis are saved permanently. Most analyses return summary data (approximately 64MB per 24 hours of audio) however some return much more data (for example, the analysis approach presented by Dong [23] generates 6GB per 24 hours of audio).

In the spirit of avoiding premature optimization [24], very little optimization is implemented initially. As algorithms become stable, performance concerns may appear through analysis of larger datasets. The optimizations to apply are chosen through profiling and greatest return for time spent. Two examples of optimizations that adhere to this principle have significantly enhanced our analysis ability: 1) segmenting of input audio files and 2) parallelization.

Long input audio files require significant amounts of RAM to processes as one block; it is not feasible to analyze input audio longer than 2 hours in duration as one block. Additionally, ecological project requirements place increasing emphasis on large-scale continuous recording – often producing files 24hrs in length. To solve this problem all analyses have been standardized on processing one-minute blocks of audio. Thus, an analysis of a 24-hour file consists of 1440 smaller one-minute analyses. Specialized programs such as *mp3splt* discussed earlier avoid sequential seeking by using indexing to allow efficient cutting of arbitrarily large audio files. The result of this optimization is effectively large scale '*streaming*' of the input audio.

A substantial side effect of segmenting input audio is that each one-minute file can be analyzed independently. A master task is responsible for creating a list of work items. Each work item cuts the audio, runs the appropriate analysis, and returns results. The master task iterates through the work items and aggregates the results. This clean separation of concerns makes it exceptionally simple to parallelize analyses and fully consume all available resources. This *intra-parallelization* dedicates one thread per logical CPU to run analysis tasks concurrently.

Although intra-parallelization sufficiently consumes the resources of most average machines, it does not fully utilize the available resources on the BigData machines. Here the ad hoc

TABLE I. SPECTRAL INDICES ANALYSIS PERFORMANCE WITH VARYING PARALLELIZATION TECHNIQUES

Machine	CPU	RAM	I/O	Analysis		Time taken[a] (m/24h)	Effective Speed up
				Threads	*Instances*		
Normal Workstation	- i5-M560 - 4 logical processors - @ 2.67Ghz each	4GB DDR3	- Hitachi HTS545025B9A300	1	1	75.05	1.00×
				8	1	41.33	1.82×
				8	>1	N/A - Unreasonable demand	
BigData	- E5-665 - 32 logical processors - @ 2.4Ghz each	256GB DDR3	- 1Gbps Ethernet - 16GB RAM cache - No local disk	1	1	74.47	1.01×
				32	1	11.61	6.46×
				32	5	3.14[b]	24.00×

a) Minutes of analysis time needed to process 24 hours of audio
b) Experiement consisted of 20 files, each 24 hours, processed in batches of 5. Total time = 62.75 minutes. 62.75 minutes ÷ 20 files = 3.14 minutes/file.

scripts that already run analyses across thousands of files (1 day of audio per file) per job were parallelized. This *inter-parallelization* runs multiple instances of the analysis process on different files. Through tuning, it was determined that each BigData machine can process five instances of an analysis executable concurrently; that is, five inter-parallelized processes, each of which has intra-parallelization enabled as well. Tuning reveals that for the BigData machines the limiting resources is CPU. The relative speed gains from inter and intra parallelization are summarized in Table 1.

D. Visualization

Visualizing acoustic data is an effective way to see details and to obtain an overview of larger datasets. Even small amounts of data are considered opaque and hard to reason about without analysis [10, 25]. Datasets that are months, even years long are common and produce numerical data that is incomprehensible. For large datasets, visualizations are increasingly becoming the only way to interpret results.

We calculate *acoustic indices* for one-minute blocks that represent content of ecological interest. Each acoustic index summarizes an aspect of the acoustic energy distribution in audio data. Three acoustic indices can be represented by different color channels. Presenting the combination of indices over time as colors in an image can expose the content of the audio and allow for navigation of audio that can be years in duration [9]. Indices can be calculated from the spectral content or waveform; there are a range of methods for calculating indices in the literature. Typical measures include SNR and amplitude. The dispersal of acoustic energy in a recording – the temporal entropy – is a promising candidate [26], as it has a good correlation with avian activity.

The choice of which three indices to combine requires measures that can be compared. We chose three indices which can easily be normalized to the range $[0, 1]$: temporal entropy, *spectral entropy* (H[s]) (a measure of acoustic energy dispersal through the spectrum) [26], and the *acoustic complexity index* (ACI), which is a measure of the average absolute fractional change in signal amplitude from one frame to the next through a recording [27]. These False-color spectrograms (see Fig 4) are built from more than one measure of the acoustic content, whereas pseudo-color spectrograms are mappings of the spectral power values to color. The combination of three indices will provide more information than a pseudo-color spectrogram if the indices used are independent.

An advantage of false-color images is that they tolerate and can even highlight data corruption and missing data. It is common to manually remove noisy or clipped recordings containing excess mechanical noise, wind, and rain, however this does not scale.

V. Work In Progress

A. Current Website Architecture

A core goal of our ecoacoustics research is to make accessing, visualizing, and analyzing large-scale acoustic data accessible to scientists. To do this we use the QCIF cloud infrastructure to host our publically accessible website. This open source application, the *bioacoustic workbench* (https://github.com/QutBioacoustics/baw-server), is designed to provide access to large-scale ecoacoustic datasets. The website successfully allows random-access to any of the ingested audio data – currently 15TB of audio.

The website provides tooling for creating *projects* and *sites* to manage audio data. From a site, access to any audio recording is possible: when loaded a visual depiction accompanies the playback of audio. Audio can be played indefinitely for radio-like listening, or can be played in sections to allow manual analysis of a segment. Annotations can be drawn on the spectrogram that, when tagged with a species name, can identify a faunal vocalization. The annotation process is useful for generating training datasets used by automated analyses [23].

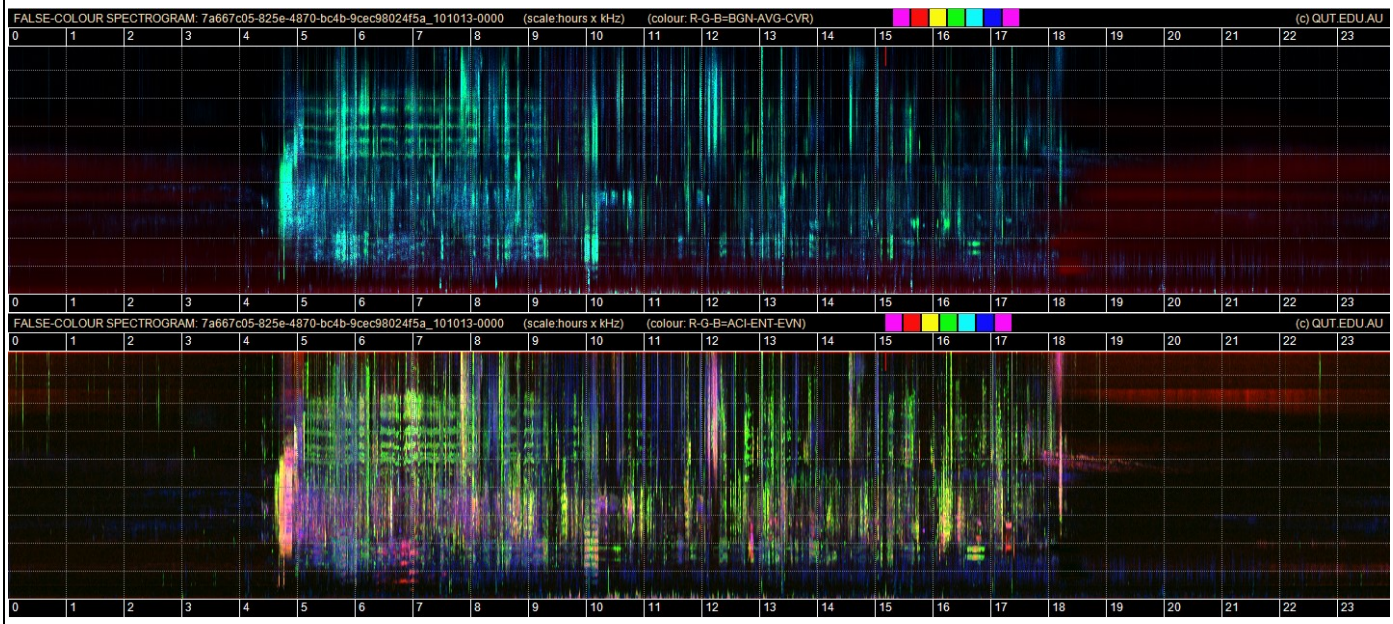


Fig. 4. Two false-color long duration spectrogram. These spectrograms use spectral indexes to visualise acoustic activity over a 24 hour period

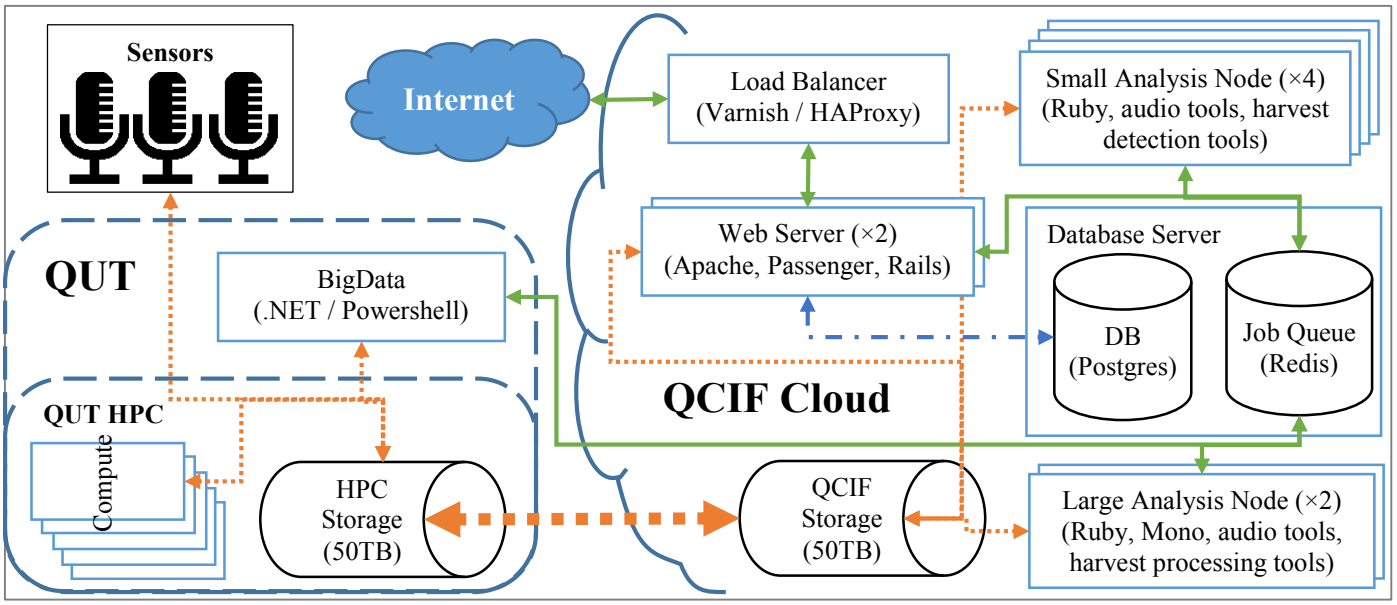


Fig. 5. Diagram of cloud scale architecture. Orange (dashed) lines represent acoustic data, green (solid) represent metadata, blue (dash-dot) represent database access

TABLE II. PLANNED ARCHITECTURE FOR SCALABLE ECOACOUSTICS WEBSITE HOSTED IN THE QCIF CLOUD

Location	VM Flavor	Instances	Resources (per instance)	Resque Queues highest priority first: `QUEUE_NAME`×Concurrency	Est. Time per Request/Job
QCIF	Web Server	2	2 VCPUs, 8GB RAM	N/A – these servers will create job items	< 2s
	Database	1		N/A – Resque host	< 1s
	Small Analysis Node	3	1 VCPU, 4GB RAM	`MEDIA`×1, `HARVEST_WATCHERS`×1	6-18s
	Large Analysis Node	2	4 VCPUs, 16GB RAM	`MEDIA`×4, `HARVEST_FILE`×4, `ANALYSIS_JOBS`×1, `MAINTENANCE`×4	1-20m
QUT	BigData Machines	1 exclusive 11 shared	32 CPUs, 256GB RAM	`ANALYSIS_JOBS`×5, `MAINTENANCE`×4	1-20m

The website is built using the Ruby on Rails framework. It utilizes our audio-tools API to cut and cache media. This provides responsive playback and on-demand loading of previously unseen segments of audio. Currently the webserver controls and executes the cutting of audio and generation of spectrograms. This is inefficient and will be extracted to separate, dedicated servers in the future.

B. *Future Architecture*

Our project has recently migrated to the QCIF cloud. The bioacoustic workbench and all audio data are currently hosted on QCIF resources; however, we have yet to fully utilize the resources available. Increased user demand and I/O strain on webservers has necessitated continued scaling. In practice, much of the analysis is driven by internal research needs and consequently run within QUT on BigData or HPC resources.

However, recent publications and increased interest in our work has resulted in progress towards more formal, scalable infrastructure. Additional functionality, including the ability to run analyses and generate false-color spectrogram images, will improve the navigation and utility of the public website.

Analysis will continue to be done locally to make use of the flexibility BigData machines afford, following the hybrid approach. We still have the need for ad-hoc scripts; however, exposing concrete analyses will improve the utility of the Bioacoustic Workbench for all users.

The job running system under development is built on Resque (https://github.com/resque/resque), a Ruby library. It uses priority queues (backed by a Redis in-memory database) to handle various asynchronous tasks. Analysis programs, audio cutting, spectrogram generation, harvesting, and maintenance jobs will be enqueued with Resque. Dedicated analysis VMs will be provisioned in the QCIF cloud to process jobs. The server architecture is shown in Fig 5 and the planned VM provisioning table and job queue distribution is shown in Table II. Additionally, Resque job runners will be installed on BigData machines to ensure compute power is never wasted – thus creating a hybrid *cloud and local* job system.

VI. CONCLUSION

Production systems for research work are difficult to provision and maintain due to the constantly changing nature of active research. The capture, analysis, and use of results from big data activities is widespread; however, practical descriptions of on-going research by groups with complex applications are needed. This paper has given an overview of the Ecoacoustic Research Group's approach to big data analysis.

The management of raw audio data, analysis programs, methods of executing programs in parallel, and resulting output is an important, significant, and time-consuming part of analyzing large data sets. It requires knowledge and experience from a range of domains implemented by a range of professionals.

Compute resources are available from a number of organizations and can provide the basis for effective big data processing. The disparate resources are often required to inter-operate. Few researchers have the background to be able to manage compute, storage, and cloud resources. As the amount of data used in the majority of disciplines increases, professional support for researchers also needs to increase.

Visualizations are an effective way to reveal patterns and summarize data that is otherwise opaque and difficult to interrogate. Developing methods for generating useful visualizations is critical to evaluating analysis algorithms. Increasing pressure to provide results from analysis of large datasets can spur researchers to remain within constraints set by professional staff; however, research requires a constant develop-and-test cycle. This tension can be addressed through freezing features and refactoring checkpoints.

ACKNOWLEDGMENTS

The authors wish to acknowledge the dedication and hard work of all members in our research group (http://www.ecosounds.org/people/people.html). In particular, we thank Jason Wimmer for conducting fieldwork and our citizen science collaborators; these are birders, conservation groups, and individuals that help analyze and verify data produced by analyses.

The authors gratefully acknowledge the funding and resources provided by Queensland Cyber Infrastructure Foundation (QCIF) and the National eResearch Collaboration Tools and Resources (NeCTAR). Grant: *QCIF NeCTAR Tools Migration Project "Acoustic WorkBench (AWB)"*.

The authors also acknowledge the resources provided by the Big Data Lab at the School of Electrical Engineering and Computer Science, QUT. Additionally we acknowledge the support and resources provided by QUT's High Performance Computing group.

REFERENCES

[1] J. Wimmer, M. Towsey, B. Planitz, I. Williamson, and P. Roe, "Analysing environmental acoustic data through collaboration and automation," *Future Generation Computer Systems,* vol. 29, pp. 560-568, 2// 2013.

[2] R. Butler, M. Servilla, S. Gage, J. Basney, V. Welch, B. Baker, *et al.*, "Cyberinfrastructure for the analysis of ecological acoustic sensor data: a use case study in grid deployment," *Cluster Computing,* vol. 10, pp. 301-310, 2007.

[3] Wildlife Acoustics. (2011, 23/05/2011). *Song Scope Product Page*. Available: http://www.wildlifeacoustics.com/songscope.php

[4] A. Gasc, J. Sueur, S. Pavoine, R. Pellens, and P. Grandcolas, "Biodiversity Sampling Using a Global Acoustic Approach: Contrasting Sites with Microendemics in New Caledonia," *PLoS ONE,* vol. 8, p. e65311, 2013.

[5] M. Towsey, S. Parsons, and J. Sueur, "Ecology and acoustics at a large scale," *Ecological Informatics,* 2014.

[6] D. Tucker, S. Gage, I. Williamson, and S. Fuller, "Linking ecological condition and the soundscape in fragmented Australian forests," *Landscape Ecology,* vol. 29, pp. 745-758, 2014/04/01 2014.

[7] I. Potamitis, S. Ntalampiras, O. Jahn, and K. Riede, "Automatic bird sound detection in long real-field recordings: Applications and tools," *Applied Acoustics,* vol. 80, pp. 1-9, 6// 2014.

[8] M. Cottman-Fields, A. Truskinger, J. Wimmer, and P. Roe, "The Adaptive Collection and Analysis of Distributed Multimedia Sensor Data," in *E-Science (e-Science), 2011 IEEE 7th International Conference on*, 2011, pp. 218-223.

[9] M. Towsey, L. Zhang, M. Cottman-Fields, J. Wimmer, J. Zhang, and P. Roe, "Visualization of Long-duration Acoustic Recordings of the Environment," *Procedia Computer Science,* vol. 29, pp. 703-712, // 2014.

[10] J. Foote, "An overview of audio information retrieval," *Multimedia Systems,* vol. 7, pp. 2-10, 1999.

[11] T. Hey. (2014, 22/07/2014). *Beyond Open Access to Open Data*. Available: http://hdl.handle.net/2142/47423

[12] E. Dumbill. (2012, 22/07/14). *What is big data? An introduction to the big data landscape*. Available: http://radar.oreilly.com/2012/01/what-is-big-data.html

[13] Y. Demchenko, P. Grosso, C. de Laat, and P. Membrey, "Addressing big data issues in scientific data infrastructure," in *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, 2013, pp. 48-55.

[14] S. Kelling, W. M. Hochachka, D. Fink, M. Riedewald, R. Caruana, G. Ballard, *et al.*, "Data-intensive science: a new paradigm for biodiversity studies," *BioScience,* vol. 59, pp. 613-620, 2009.

[15] G. Bell, J. Gray, and A. Szalay, "Petascale computational systems," *Computer,* vol. 39, pp. 110-112, 2006.

[16] Xeno-canto Foundation. (2014, 22/07/14). *Colophon and Credits*. Available: http://www.xeno-canto.org/about/credits

[17] S. Duan, J. Zhang, P. Roe, J. Wimmer, X. Dong, A. Truskinger, *et al.*, "Timed Probabilistic Automaton: A Bridge between Raven and Song Scope for Automatic Species Recognition," in *Twenty-Fifth IAAI Conference*, 2013.

[18] L. J. Villanueva-Rivera and B. C. Pijanowski, "Pumilio: A Web-Based Management System for Ecological Recordings," *Bulletin of the Ecological Society of America,* vol. 93, pp. 71-81, 2012/01/01 2012.

[19] L. Dayton, "Giant telescope to 'create hundreds of jobs'," in *The Australian*, ed, 2012.

[20] P. J. Guo and D. R. Engler, "Towards Practical Incremental Recomputation for Scientists: An Implementation for the Python Language," in *TaPP*, 2010.

[21] S. R. Kohn, G. Kumfert, J. F. Painter, and C. J. Ribbens, "Divorcing Language Dependencies from a Scientific Software Library," in *PPSC*, 2001.

[22] T. Hey, S. Tansley, and K. M. Tolle, "Jim Gray on eScience: a transformed scientific method," in *The Fourth Paradigm: Data-Intensive Scientific Discovery*, ed Redmond, Washington: Microsoft Corporation, 2009.

[23] X. Dong, M. Towsey, Z. Jinglan, J. Banks, and P. Roe, "A Novel Representation of Bioacoustic Events for Content-Based Search in Field Audio Data," in *Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on*, 2013, pp. 1-6.

[24] D. E. Knuth, "Structured Programming with go to Statements," *ACM Computing Surveys (CSUR),* vol. 6, pp. 261-301, 1974.

[25] M. Towsey, J. Wimmer, I. Williamson, and P. Roe, "The use of acoustic indices to determine avian species richness in audio-recordings of the environment," *Ecological Informatics,* vol. 21, pp. 110-119, 5// 2014.

[26] J. Sueur, S. Pavoine, O. Hamerlynck, and S. Duvail, "Rapid Acoustic Survey for Biodiversity Appraisal," *PLoS ONE,* vol. 3, p. e4065, 2008.

[27] N. Pieretti, A. Farina, and D. Morri, "A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI)." *Ecological Indices,* vol. 11, pp. 868–873, 2011.

Chapter 4
Rapid Scanning of Spectrograms

4.1 Introduction

The publication in this chapter describes the research conducted for testing the rapid scanning methodology. In the annotation system studied by this thesis (see Chapter 3), the annotation stages of event *detection*, *segmentation*, and *classification* occurred simultaneously. It was hypothesised separating event detection and classification tasks could improve annotation efficiency. User interfaces that focus on one task have been shown to perform better (Reeves et al., 2004); better tooling and simpler tasks allow users to become more efficient.

This research focusses on a method designed for identifying certain types of acoustic events in a rapid manner. The approach, labelled *rapid scanning of spectrograms*, shows the visualisation of the audio data (as a spectrogram) to a participant in a rapidly sped-up sequence with the audio disabled. This technique has the potential to speed up analysis by a factor of 12× for the detection of certain types of calls, under certain conditions. This method of analysis has a theoretical upper bound of a factor of 24× improvement (0.5s exposure); at this speed, the limits of human reaction and decision times are reached.

This concept models emergent behaviours of participants and incorporates the behaviour into analysis tooling. By automating a small part of the tooling, it is possible to allow a participant to quickly identify the presence (or not) of interesting vocalisations over hours of data. Once this process has finished, participants can then be sent back to only those sections where acoustic events were found to fully annotate the data. This chapter specifically addresses sub research question 1 (section 1.2): *Can the faunal event detection speed of analysts be enhanced?*

The publication details and demonstrates the rapid-scanning technique as well as the experiment that was conducted in order to measure the effectiveness of this method. The research was conducted by creating a UI Prototype (used by participants) that was instrumented by quantitative and qualitative protocols. The experiment conducted was a fixed study, with quantitative and qualitative measures. Part of the research outcome was to produce a feasible UI artefact. The quantitative part of the experiment was conducted to measure the accuracy of participants using the interface at different speeds. The qualitative survey was conducted to capture user opinions of the system. The survey was conducted on QUT's instance of *Key Survey* – A web-based survey platform.

Participants were contacted via social media, personal communication (email, phone contact), and flyer distribution. In total, 73 participants external to QUT participated in the project. Ethics considerations for this chapter fall under the low risk human ethics policy detailed in section 1.6.

The study found that fast-forwarding spectrograms past a participant at a rate of 12× normal speed (2.0 second exposure of a 24 second static spectrogram), resulted in a trade-off in accuracy (of 17%). This method has potential for future testing and integration with production interfaces.

4.2 Conference Paper – Rapid Scanning of Spectrograms for Efficient Identification of Bioacoustic Events in Big Data

Truskinger, A., Cottman-Fields, M., Johnson, D., & Roe, P. (2013). *Rapid Scanning of Spectrograms for Efficient Identification of Bioacoustic Events in Big Data*. Paper presented at the 2013 IEEE 9th International Conference on eScience (eScience), Beijing, China.
http://dx.doi.org/10.1109/eScience.2013.25

This conference paper has been peer reviewed and published.

4.3 Statement of Contribution

RESEARCH STUDENTS CENTRE
Examination Enquiries: 07 3138 1839
Email: research.examination@qut.edu.au

Statement of Contribution of Co-Authors for Thesis by Published Paper

The authors listed below have certified* that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter:

Publication title and date of publication or status:

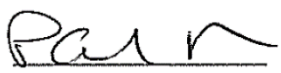
Rapid Scanning of Spectrograms for Efficient Identification of Bioacoustic Events in Big Data

Published, 2013

Contributor	Statement of contribution*
Anthony Truskinger	Wrote the manuscript, designed the experiment, created the software prototype
Signature 	
Date 21/05/2015	
Mark Cottman-Fields*	Helped build the software platform for the experimental interface.
Daniel Johnson*	Aided in experimental design and statistical analysis
Paul Roe*	Supervisor – Oversaw and contributed to the entire paper

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Paul Roe  20/5/15

Name Signature Date

Rapid Scanning of Spectrograms for Efficient Identification of Bioacoustic Events in Big Data

Anthony Truskinger, Mark Cottman-Fields, Daniel Johnson, Paul Roe
Computer Human Interaction / Computer Science, Science and Engineering Faculty
Queensland University of Technology
Brisbane, Australia
{anthony.truskinger, m.cottman-fields}@student.qut.edu.au, {dm.johnson, p.roe}@qut.edu.au

***Abstract* — Acoustic sensing is a promising approach to scaling faunal biodiversity monitoring. Scaling the analysis of audio collected by acoustic sensors is a big data problem. Standard approaches for dealing with big acoustic data include automated recognition and crowd based analysis. Automatic methods are fast at processing but hard to rigorously design, whilst manual methods are accurate but slow at processing. In particular, manual methods of acoustic data analysis are constrained by a 1:1 time relationship between the data and its analysts. This constraint is the inherent need to listen to the audio data. This paper demonstrates how the efficiency of crowd sourced sound analysis can be increased by an order of magnitude through the visual inspection of audio visualized as spectrograms. Experimental data suggests that an analysis speedup of 12× is obtainable for suitable types of acoustic analysis, given that only spectrograms are shown.**

Keywords—sensors; acoustic data; spectrograms; big data; big data analysis; crowdsourcing; fast forward

I. INTRODUCTION

Acoustic sensors provide an effective way to scale biodiversity monitoring to large scales [1-3]. Acoustic sensors record large amounts of data continuously and objectively over extended periods. There are many ways to analyze these huge datasets, ranging from completely manual approaches, to fully automated methods of detection.

Automated vocalization detection and classification of fauna in recordings has been the subject of much research. There are many examples of single species detectors [3-5], fewer algorithms capable of detecting multiple species [6], and some examples of general purpose tools capable of general audio data analysis [7-9].

However, automated methods of analysis are not perfect. They can suffer high rates of false positives and false negatives [10, 11] and are time-consuming and expensive to develop. Extracting good training sets is particularly time consuming, requiring extensive tuning and adaptation for different environments [10, 12].

An alternative approach to automated methods is to use crowd-based methods of analysis. The idea is that it is possible to outsource a complex classification task to a crowd of interested participants. In these scenarios, technology can be used to assist with the menial parts of the analysis tasks. We term this combination of manual and automated approaches as semi-automated analysis. Varying levels of automation and human participation result in a spectrum of methodologies that exist between the two extremes.

In our research project, we use a semi-automated analysis methodology in addition to developing fully automated methods of detection [3]. Currently, in our semi-automated system, participants analyze data in a web interface by playing back audio collected from sensors. The audio is played, along with a visual representation of the sound displayed at the same time. This visualization is a spectrogram – a time/frequency graph that can show the 'shape' and intensity of the underlying audio. The spectrogram is currently translated left (animated horizontally to screen left) at a speed that is equivalent to the audio playing (approximately 45px/s). We label this speed as real-time (or 1×). Fig. 1 shows a screenshot of this software.

The large amount of audio data that needs to be analyzed places strain on the limited resources of our volunteer participants. As we observed our participants analyzing data, a unique behavior was noticed when participants were trying to identify only one species at a time. They would rapidly 'scan' through each section of audio that was loaded into our online analysis tool. This scanning involved waiting for each 6 minute block to load (~3MB of audio, 1MB of images), dragging the seek/progress/navigation bar from start to end at a speed they were comfortable with, stopping only when they found their target pattern. Accordingly, without listening to the audio and by relying on the spectrograms alone to identify their target vocalisations, a participant could process the 6 minute block in seconds. This ad hoc method is suboptimal due to the loading of redundant data and the limiting size of audio segmants that can be loaded at any one time.

To optimize the process and determine the degree of accuracy that can be achieved, this paper tests this ad hoc 'rapid scanning' method of semi-automated analysis for viability.

A. The challenges of big acoustic datasets

Our project's acoustic sensors collect on average eight days' worth of audio data every day. Meanwhile, semi-automated analysis currently takes a participant approximately two hours to analyze an hours' worth of data, at reduced resolution [13]. Hence, if we wish to scale our audio analysis, an increase of efficiency in the analysis process is required.

One limiting factor is the consumption of audio data. Audio data is ideally consumed in real-time (1× speed). Other speeds distort the sound resulting in a different interpretation of the original sound by a human. However, a spectrogram, since it is

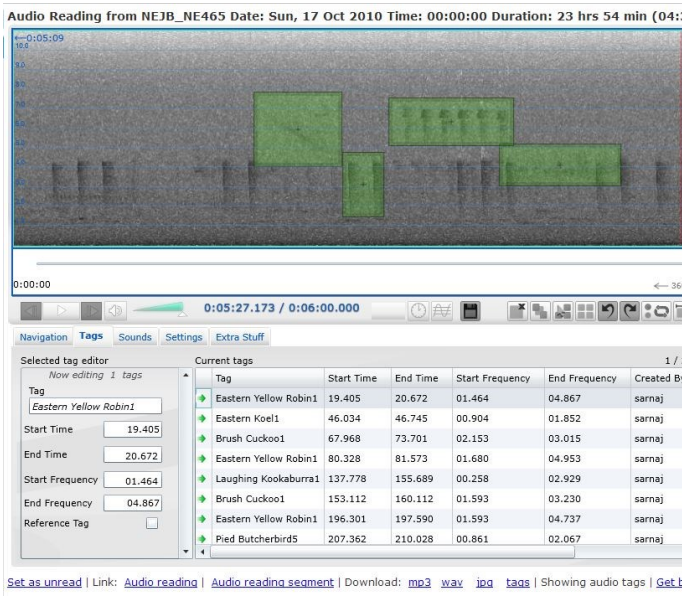


Fig. 1. A screenshot of out current annotation software

only an image, remains visually identical no matter what speed it is translated at. The limiting factor is the amount a participant can perceive in an image with limited temporal exposure.

By disabling the audio and speeding up (fast-forwarding) the animation of the spectrograms, there is the potential to have our participants analyze the data faster, without a severe loss of accuracy. This paper presents an experiment that tests the aforementioned concepts for feasibility. If feasible, this paper will add another method for semi-automated data analysis to the existing toolbox of techniques.

II. Related Work

A. Related Citizen Science Work

Galaxy Zoo is an example of a successful citizen science project that utilizes a crowd sourced image classification model – similar to the model we employ for identifying patterns in spectrograms. The Galaxy Zoo project uses their volunteers to classify the morphology of galaxies from the Sloan Digital Sky Survey by showing them images of the galaxies and asking them to pick a similar shape [14]. Importantly, participants complete the tasks at their own pace and classification speed is not emphasized. Instead, Galaxy Zoo scales out their analysis by gathering large numbers of active participants. A focus on faster classification times would not work well with the current versions of Galaxy Zoo, as the classification task asks multiple questions about each presented image.

WhaleFM [15] is a derivative of the Zooniverse project which operates Galaxy Zoo. Again, the core concept is to harness the collective intelligence of volunteer participants to analyze images. However, WhaleFM differs in that it shows spectrograms of whale song to participants for classification into one of several classes. This is very similar to this papers' stated task. Whales create vocalizations on the lower end of the spectrum of human hearing, thus, it is not always easy to hear them. By visualizing the sound with a spectrogram image, it lets the participants match the image in their own time, not constrained to real time audio. The WhaleFM paper by Saigh *et al.* [15] shuffled the order of the spectrograms shown to the volunteers used in the paper's experiment. The paper did not reveal how long it took its participants to classify the whale song patterns. Like Galaxy Zoo, it has multiple possible classifications for vocalizations, making it potentially difficult to scale in speed.

A paper by Lin *et al.* [16] demonstrated a similar rapid-analysis technique. The paper uses human participants to detect acoustic events of interest in spectrograms. The user can jump to any point in the audio stream and adjust the zoom of the visualization at the same time. Their study was conducted in order to bypass the time constraints of listening to and analyzing audio data. Additionally, they found that spectrograms were a good choice for visualizing their data because even untrained participants were capable of completing their assigned tasks of locating acoustic events. Participants were given 8-minute blocks of time to identify as much content as possible in 80 minutes of audio. The spectrograms are enhanced and shown in a zooming-style interface that allows participants to control the scale of the spectrograms (and thus the audio) that is shown. When identifying an event the user has the option to playback the associated audio. In practice, the authors stated acceptable results with their 10× speed increase. Importantly, their experiment was unstructured – participants chose where and when they stopped and listened to audio data.

B. Perception and Reaction times

The widely accepted minimum reaction time for visual stimuli in humans is about 200ms [17, 18]. However, reaction time slows when a choice needs to be made, as when classifying something, reaching 400ms and higher depending on the complexity of the image [17].

Biederman [19] states that image processing in humans is component based. This means humans are good at looking for shapes in images, like the sort of shapes often seen in spectrograms. The paper also states that as the number of components presented increases, error goes up. Biederman suggests that at least one second is required for the analysis of a degraded image. A degraded image is defined as one missing parts, like contours, surfaces, or other gaps. Spectrograms can be complex and vocalizations within can often be missing components.

Konishi *et al*. [20] did a study on brain activity for a go/no-go task. They trained participants to respond within a 300ms reaction time to a go/no-go task (press a button for positive or another negative) for a simple visual stimulus.

Joubert *et al.* [21] have done several studies of times taken for participants to classify a scene. Generally, they flash an image up for a very short amount of time (20ms) and see categorization into one of two groups (i.e. go/no-go) in around 400ms.

In summary, the best reaction times cannot be less than 200ms for a classification task and an average of around 400-600ms is expected for classification of an image like a spectrogram.

III. Experiment Design

This experiment should assess the viability of the rapid scanning methodology through the construction a new and appropriate interface. To measure the net data processing speed, the test interface will show different speeds and measure which settings result in the best analysis. Ideally, the experiment should attempt to understand how the rapid scanning methodology would scale. The experiment must also be web browser compatible. Our existing analysis systems runs in an online environment and it would be ideal to integrate the work if it were feasible.

A small survey will also be issued to participants after they complete the experiment.

A. Limitations

1) Soundscape

Vocalizations of interest must be easy to identify by human participants. This means that the vocalization should be distinct and likely to occur in moderately empty audio signals. When working with relatively empty audio signals, it is still possible to have a complex and dynamic acoustic profile in the recordings. This variation is caused by a variety of non-target acoustic features such as rain, wind, crickets, or complex non-bioacoustic events. When combined these artifacts can prevent simple automated detection techniques from working effectively.

The human component of the rapid scan methodology is what makes this idea feasible. A human participant can intelligently distinguish between infrequent faunal vocalizations and sudden intense or complex periods of uninteresting audio. However, humans have limits of perception and focus. Analyzing with the intent of classifying every species present at once, or analyzing in dense areas of bioacoustic events will overwhelm a human participant – especially when asked to do so quickly. Thus, the rapid scan method is thought to be most useful for speeding up the analysis of the sparse, time-consuming, night section of an acoustic day.

2) Participants' tasks

Typically when analyzing audio data, participants are tasked with annotating vocalizations. Each annotation action involves drawing a bounding box around the portion of the spectrogram containing the vocalization and then associating one or more textual tags with said bounding box. These annotations form the core data output for this research project; however, they are also time consuming to create. The rapid scan methodology is intended to analyze data rapidly. If a participant were to stop every time they detected a vocalization and then annotate it, the desired speed up in analysis would likely not be obtained. Instead of full annotations, a simpler method of detection was chosen: a simple positive 'hit' button.

Once points of interest are discovered (as hits), it is then possible to get any participant to return to the data later to properly annotate. The rapid scan process still provides a service by filtering out the large sections of audio that contain no interesting vocalizations. In other words, this is filtering with human vision to break up a time-consuming task into components of work.

3) Inclusion of a 'negative' answer

Ideally, there would only be a positive hit answer in the user interface as it is all that is needed to complete the rapid scan task. Experimentally, this would mean it is not possible to determine the difference between a participant failing to respond and a negative response. Thus, a negative response option was included to enable this information to be gathered.

4) Disabled audio playback

Enabling playback of audio for the rapid scan methodology was considered. It would be ideal for participants to hear the audio data – it is a powerful discriminator for distinguishing between signal and noise. Audio also helps explain spectral components in the spectrogram and helps to keep the task interesting for participants. However, playback of audio is constrained to a 1× speed – this is the very speed constraint the rapid scan methodology is trying to avoid. Any playback of audio would reduce the effectiveness of the rapid scan methodology.

B. Hypothesis

Research question: by manipulating the animation speed of the spectrograms, to make them display faster, will participants be able to detect interesting acoustic events at an increased speed, with an acceptable trade off in accuracy.

The null hypothesis (H_0) for this experiment is: No difference in accuracy will occur at different exposure speeds. The alternative hypothesis (H_1) for this experiment is: That accuracy will be effected by speed of presentation such that accuracy will decrease at higher speeds.

C. Experimental Interface Design

Flashcards were chosen over the project's traditional animated image translation for simplicity. A flashcard is simply a card that shows information – they are often learned for memorization tasks. We use the term flashcard in a digital sense to refer to a series of spectrogram images that are to be flashed past an analyzer-participant. Flashcards are simpler than a translation animation; they simply need to be shown for some duration and then hidden again. This means they do not move distractingly during viewing, allowing participants to scan according to their personal preference rather than forcing them to scan left to right. For a traditional translated image approach, it is required to animate not just one image but neighboring off-screen images as well, in a demanding animation loop. A translating image approach requires a concept of scale (pixels per second) and is inherently limited by the rendering capabilities of the browser (often 60fps).

The amount of audio data shown with each flashcard was set to 24 seconds. This amount was chosen because a 24-second spectrogram, at standard scale (≈43px/s) fits well within most screen resolutions; it is 1033px wide by 256px high. The spectrograms are created with a 512 sample window and no overlap. The duration of 24 seconds also divides conveniently into 120 seconds – many of the smaller recordings available are two-minute long blocks of audio data.

A screenshot in Fig. 2. shows the instructions page that was given to each participant between each segment of analysis. When presented to participants animations emphasized core

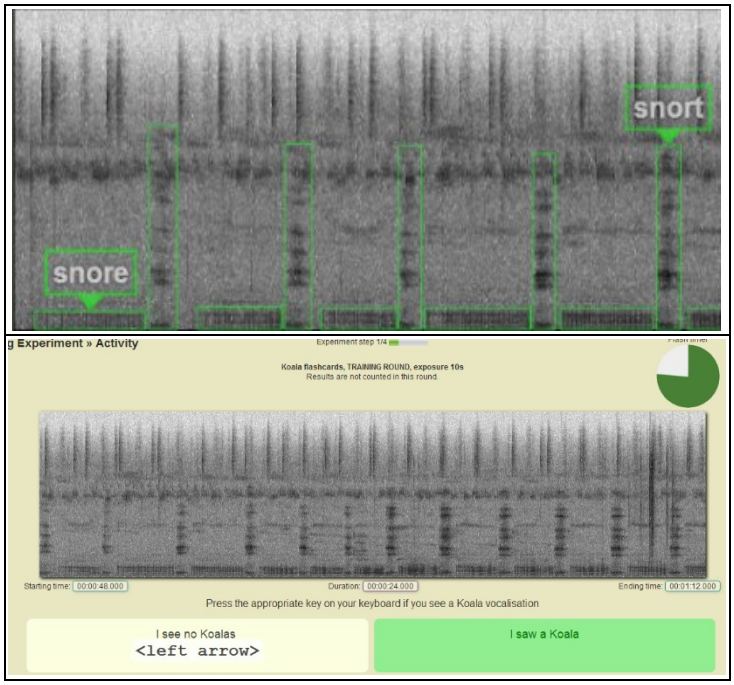


Fig. 2. Screenshots of the experiment interface. Top: The spectrogram from the training page. Bottom: An example "yes" hit (a true positive) on the classification page.

components of the instructions. In particular, the outlining in bright green of the example vocalizations was animated and labelled. Additionally, the exposure speed, number of flashcards in the segment, and the key bindings were bolded to make them stand out.

The classification page (Fig. 2) consisted of timestamps (the bounds of the flashcard), an exposure countdown timer, a pause / resume button, and a segment progress bar. A lead-in countdown appears on the classification page; it instructs users to place their hands on the keyboard and displays a ten second countdown to ready the participant before each segment of the experiment started.

D. Experiment protocol

The experiment was conducted according to the following protocol:

1. The experiment was advertised and participants were contacted via email, in person, and through social media.
2. The landing page was the first thing participants saw. On this page, the basic details of the experiments were shown. Participants were encouraged to read the ethics statement and were required to consent to their participation in the experiment.
3. A segment order protocol was created for the participant. See the following section 'Segment Order and Randomization Protocol' for more information.
4. The training round was conducted: each flashcard lasts 10 seconds and only three flashcards are shown.
5. The main experiment was then run. Three rounds were conducted using the datasets and the speed combinations defined by the segment order protocol. These rounds showed a total of 165 flashcards.

TABLE I. SPEEDS TESTED IN EXPERIMENT

Speed	Exposure time	Rate = $\frac{exposure\ time}{24s}$
Slowest	5.0s	4.8×
Medium	2.0s	12.0×
Fastest	1.0s	24.0×

6. End of experiment: the end screen shown and survey link were displayed and the data was sent back to the server.
7. Survey optionally completed by participant

E. Speeds

The flashcard exposure speeds tested in this experiment are shown in Table 1. A range of speeds were chosen around the 2s exposure mark. The 2s mark was chosen based on observations of the ad-hoc rapid scan methodology. The data used in the experiment was annotated previously and thus a real-time speed was not included as a control. The real-time data was used as the baseline accuracy measure.

F. Datasets

The data chosen for this experiment was taken from a project that deployed sensors located at St Bees Island, Queensland Australia (latitude: -20.914, longitude: 149.442). This island has a population of Koalas (*Phascolarctos cinereus*) relatively unaffected by mainland Australia, making it a source of interesting research [22]. Koala vocalizations were chosen because it is known that Koala usually call at night [23]. Koala vocalizations are also easy to distinguish and identify – they are long, loud, and distinct.

Data was taken from two different sites at St Bees, from 30/September/2009 to 16/August/2011. Recording timestamps spanned from 17:00 through to 04:30. The sensors used were 3G phones that recorded 2 minutes of audio every half hour.

For the experiment, three datasets, one for each speed, totaling 66 minutes of data (22 minutes for each dataset) were chosen. The idea was to provide enough data for each participant to complete, in order to simulate what the experimental task might be like at a large scale, balanced against the time constraints of the participants.

The recordings were included in their entirety, unedited, into the dataset when a Koala Bellow was found. All sections chosen were previously annotated so that reference data was available. There were an unnatural number of positive hits in the experiment datasets. In a real world example, fewer recordings would have a Koala vocalization present. This experiment was designed so that the presence of Koala vocalizations occurs approximately 50% of the time. In actuality, vocalizations occur in 40% of the flashcards.

G. Segment Order and Randomization Protocol

In the experiment, it was desirable that each speed was tested on each dataset.

If all the participants experienced the varying speed tests in order, (i.e. 5s, 2s, 1s) they might have been unfairly trained for the faster speeds. To avoid a training bias, the combination of

speeds was set to be *order important*. Thus, the three speeds produce six permutations.

Given three datasets, there were 18 possible combinations for the experiment. The combinations were tracked and handed out evenly to each new participant of the experiment. This ensures that the participants completed roughly the same number of each the possible segment orders.

In each dataset, the order of the flashcards that are shown were randomized. This ensured that participants were very unlikely to receive either a) contiguous flashcards or b) an order of flashcards that might unfairly bias them (e.g. due to unintentional training).

IV. Results

An error in storing the data on the experiment server rendered some of the experimental results unusable. This created a disparity between the number of survey results and the number of experimental results. There were 46 experimental results and 73 survey responses. The corrupted experimental data was discarded and the remaining experimental results verified for integrity. Since the survey responses are independent of the experiment data, all of the survey responses were used.

A. Main experiment overview

A script for data manipulation processed the JSON files sent back to our server from the website. The data was then subsequently analyzed by Microsoft Excel 2013 and verified by IBM's SPSS Version 21.

Experimental results were collected from 2/April/2013 through to 21/April/2013. In total 73 experimental results were collected. Twenty-seven experimental results were deleted due to data corruption, leaving 46 valid responses. All subsequent reports on experimental data will include only the data from the 46 valid experimental results.

Throughout the experiment period 7 728 flashcards were shown generating 8 023 hits, where a hit is a decision made about a flashcard. Changes in decision were possible meaning on average there was 1.04 hits per flashcard. A miss was a flashcard that did not receive any hit. Misses occurred in 512 (7% of) flashcards.

Given that each flashcard showed 24 seconds of audio data, 51.52 hours' worth of data was analyzed during the experiment. This analysis was completed with 7.02 hours of human effort; this included, training time, pauses, breaks between segments, and download time. Without pauses included, only minimal time was spent downloading spectrograms and reading the instructions for each segment. The human effort spent without pause breaks of 6.01 hours, computed to an effective average exposure speed of 2.80s/flashcard (8.6×, average across all speeds, including training). The expected average exposure time across all flashcards was 2.55s/flashcard (9.4×).

On average, each segment order was completed 2.56 times.

TABLE II. SEGMENTS BREAKDOWN FOR EXPERIMENT RESPONSES

Speed (s)	training	DS1	DS2	DS3	SUM
10	46	0	0	0	**46**
5	0	14	17	15	**46**
2	0	15	15	16	**46**
1	0	17	14	15	**46**

TABLE III. MARKING STYLE

	Positive (non-ambiguous)	Negative (ambiguous or non-existent)
True	TP	TN
False	FP	FN
Miss	MP	MN

B. Main experiment results breakdown

This section reports participants' accuracies at different speeds. Accuracy is the statistic we used for summarizing responses to flashcards. Accuracy is defined as:

$$a = \frac{TP + TN}{P + N} \quad (1)$$

where a positive or negative was determined by the presence of a koala vocalization and a true or false was determined by marking a participant's answer against the relevant flashcard. Accuracy was chosen because it represented the statistic we were most interested in and because it was not defined by *false* cases. This is useful because there were two types of *false* cases: an incorrect decision and a *miss* – where a participant has failed to respond within the exposure time.

a) Consistency of Datasets

As described, three datasets were created for use in the study. These datasets were then presented to participants at various speeds. These datasets were presented with their spectrograms randomly shuffled. Before testing the performance of participants at different speeds, it is important to confirm that no difference in accuracy was found between datasets (as this would indicate a confound resulting from the random allocation of spectrograms to each dataset). To ensure no systematic error was unintentionally introduced into the study in the form of datasets that were more or less difficult to analyze, regardless of speed, inferential statistics were used to confirm that all datasets were equivalent.

There were ten outliers in the data as assessed by inspection of boxplots. In addition, accuracy was not normally distributed for each dataset as assessed by Shapiro-Wilk's test ($p < 0.001$). Thus, an ANOVA was not a suitable test since its assumptions were not met. Instead, a Kruskal-Wallis test was run to determine if there were differences in accuracy for flashcards between datasets.

Initially, the datasets were collapsed across speed and compared. No statistically significant differences were found between the three datasets, $\chi^2(3) = 5.638$, $p = 0.131$, indicating that no dataset was more or less difficult to analyze than any other. Because a slightly different proportion of each dataset was used at each speed due to the final number of participants

TABLE IV. DATA SET BREAKDOWN

Dataset	Instances	Accuracy (mean)	SD	Miss Rate (%)
training	138	0.80	0.26	0.06
DS1	2530	0.80	0.16	0.06
DS2	2530	0.79	0.19	0.07
DS3	2530	0.82	0.19	0.07

in the study, further Kruskal-Wallis tests were done between the datasets for each speed individually. These tests also revealed no significant difference between the datasets. In sum, as required to allow for a valid test of performance at difference speeds (see below), no difference in difficulty of datasets (participant performance) was found between the three datasets.

b) Effects of speed on accuracy

To determine the effect of exposure speed on accuracy and test hypothesis 1, a series of inferential tests were conducted. A repeated measures ANOVA was conducted to determine whether there were statistically significant differences in Accuracy over varying flashcard exposure speeds.

There were two outliers in the data as assessed by inspection of boxplots. One outlier (accuracy = 0.16) occurred at speed 2 where the user had stopped responding. The other outlier (accuracy = 0.44) occurred at speed 5, where it seems the participant got the positive hit and negative hit responses mixed up. Both participants were removed from the dataset. To assess the assumption of normality, skewness and kurtosis values were calculated at each speed. All variables were found to be skewed. The data was then transformed with an arcsine ($\sin^{-1}$) transformation. Skewness and Kurtosis values were then recalculated and found to be acceptable for all variables. To allow for the violation of the assumption of normality, all analyses were conducted with both the transformed and the non-transformed variables. No differences were found in the pattern of results, so for ease of interpretation the results with the non-transformed variables are reported below.

Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated, χ^2 (2) = 35.125, $p < 0.001$. Therefore, a Greenhouse-Geisser correction is applied ($\boldsymbol{\varepsilon}$ = 0.736). Accuracy was statistically significantly different at the different speeds during the experiment, $F(1.277, 54.893)$ = 16.864, $p < 0.001$, partial $\eta^2 = 0.282$. Accuracy decreased from 5s (0.87 ± 0.01), to 2s (0.85 ± 0.18), to 1s (0.73 ± 0.21), in that order. *Post-hoc* analysis with a Bonferroni adjustment revealed that accuracy statistically significantly decreased from the 2s speed to the 1s speed (0.12 (95% CI, 0.201 to 0.042), $p = 0.001$). Additionally, accuracy statistically significantly dropped from the 5s speed to the 1s speed (0.15, (95% CI, 0.069 to 0.227), $p < 0.001$). However, there was no significant increase in accuracy from the 2s speed to the 5s speed (0.03, (95% CI, -0.070 to 0.060), $p = 0.170$).

TABLE V. SPEED BREAKDOWN

Speed (s)	Instances	Accuracy (Mean)	SD	Miss Rate (%)
10	138	0.80	0.26	0.06
5	2530	0.87	0.09	0.01
2	2530	0.83	0.16	0.05
1	2530	0.71	0.22	0.14

c) Summary

The results from the repeated measures ANOVA allowed us to reject the null hypothesis that accuracy is the same across all speeds. Furthermore, operating at speed 2 produces an accuracy that is not significantly different than operating at speed 5; thus accuracy is kept with the faster 2s speed. However, working at the fastest speed (1s) resulted in a significant drop in accuracy in comparison to working at the slower speeds.

2) Hit distributions

Every hit (classification) event of a flashcard was recorded with the event's timestamp. These hits were compared between the different speeds in Fig. 3.

When analyzing the hit timestamps some inconsistencies were noticed with the timestamp data. Investigation into these inconsistencies suggested that some form of lag spikes or pauses intermittently affected the timestamp calculation. In total 200 hit instances were excluded from the 8023 instance hit dataset because they fell outside the logical bounds of the exposure period for their associated flashcards.

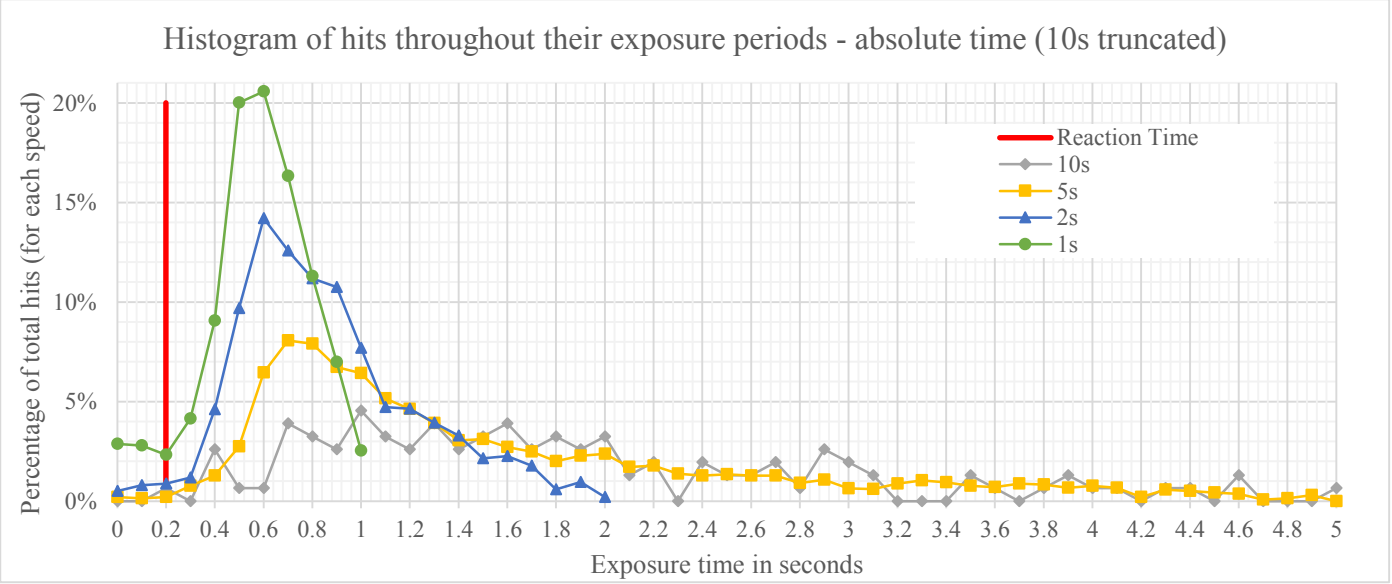


Fig. 3. Histogram of hit distributions with an absolute time x-axis, broken into 0.1s bins. The y-axis is normlized as percentage of hits within each speed

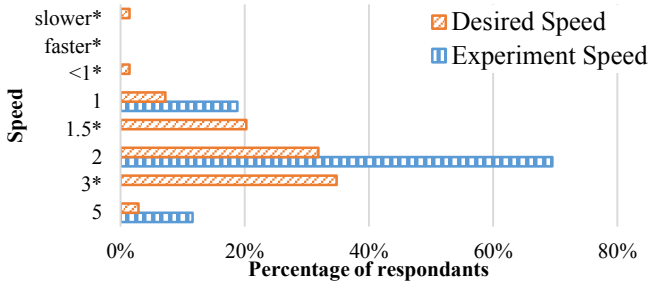


Fig. 4. Preferred speed from the experiment and desired speed for long amounts of work

TABLE VI. AGE AND VOCATION BREAKDOWN

Age	Percentage
<18	0.00%
18 - 25	41.10%
25 - 35	24.66%
35 - 45	12.33%
45 - 55	9.59%
55 - 65	9.59%
65+	2.74%

Qualifications	Response
High School Diploma	28.77%
TAFE Diploma	15.07%
Graduate Degree	28.77%
Post graduate degree	24.66%
Post-doctoral qualifications	2.74%

C. *Survey*

The survey received 76 responses, 56% male, 43% female, and 1% other.

Half of the participants (52%) had never gone looking for wildlife recreationally. The rest were Birders (15%), bush walkers (31%), and snorkelers/divers (16%), with 2 responses from herpetologists, and 2 people that lived on farms. When asked about the years of experience they had doing recreational biology, 45% responded with 'No amount of time'. Twenty-four participants had experience greater than 5 years. Two professional biologists participated.

The provided instructions were adequate for 92% of participants. Two people commented on whether both parts of training pattern had to be included for the pattern to be considered valid. Participants commonly asked for more example training images.

Almost all of the participants (70%) preferred a 2 second exposure time out of the speeds they completed in the experiment. When asked about other speeds they would prefer, participants liked speeds 3, 2, 1.5 with 35%, 32%, and 20% respectively (Fig. 4.). Four participants advocated a variable speed.

Other comments included requests for bigger spectrograms and more training samples that included answers. Participants found the 5s speed boring and uninteresting. Participants also reported feeling stressed, uncomfortable, and frustrated during the 1s speed. Most agreed that the 1s speed was too fast. At least one participant gave up answering negative hits. Two participants wanted to progress through the flashcards at their own pace. Generally, participants wanted to listen to the audio.

V. DISCUSSION

This paper's research question seeks to determine if it is viable to flash images of audio past participants at high speeds for analysis. Before answering the question of viability, it is necessary to determine the speed that performed best. The best speed can then be used to determine viability.

A. *The best speed*

The main experiment quantitatively showed that of the three speeds tested, there was a significant drop in accuracy for only the fastest speed, 1s (24×), when compared to the other speeds.

For the 2s (12×) & 5s (4.8×) speeds there was no significant difference in accuracy among the participants. This means there is no significant drawback in accuracy for tasking participants to operate at a 2s speed over 5s.

Additionally, the miss rate for participants at speeds 2s and 5s were 5% and 1% respectively. For the 1s speed, this rate jumped to 14%. These misses were partially explained by a single participant that declined to answer for negative flashcards for the 1s speed only – that accounted for 0.8% of flashcards.

B. *Hit distributions*

The hit distribution data (Fig. 3) provides insight into when in the exposure period users were responding.

The median responses for all three experimental speeds were between 500ms and 800ms – approximately what the literature suggested it should be. As the speed increased the median hit time decreases. We speculate that the increased speed is forcing the participants to lower their average reaction time.

When the tails of hit distributions were compared, we see that for speed 1, at its upper bound, the histogram shows a non-zero value of ~2.5%. This meant that on average a group of participants was not responding within the time constraint. For the 2s and 5s speeds, the histograms demonstrate a more relaxed tail of diminishing responses. The last data point for 2s is at 0.002% of hits and the 5s speed actually hits zero – indicating all users had finished responding within the time constraints. The difference in completion of hits within the allotted time between 2s and 5s was negligible (0.002%). This means, that the extra three seconds of time between the speeds is wasted, given all responses can be accounted for without the extra time.

Finally, the red line on the hit distribution graph represents the *upper* bound of human reaction time performance (200ms). Discussed in the related work section, it is extremely unlikely to see a legitimate response within the first 200ms of exposure of a flashcard. We speculate any hits that occur within this 200ms period are invalid, either caused by panic, or delayed reactions – i.e. a participant decided how to classify a card too late and accidently responded to the next flashcard in the sequence. With the 2s and 5s speeds only a very small percentage of hits occurred within this first 200ms; 0.5% for 5s, and 2% for 2s (cumulative where $t < 200$ms). However, for the 1s speed, the cumulative hits reached 8% ($t < 200$ms). This means, of the 2530 flashcards shown there were 202 responses for which it is impossible for them to be legitimate.

Tying the miss rate (14%) in with the impossible-response-time rate (8%) for the 1s speed, there's a minimum 20% error rate for speed 1 – much higher than speeds 5 or 2 (1.5% & 7% minimum error rates respectively).

C. *Survey data*

The qualitative data received from the survey produced a wide range of results. Gender was roughly split, age was

skewed towards the 18-25 bracket, and vocation was roughly evenly distributed. Importantly, 45% of respondents indicated they do not recreationally look for wildlife. This means a reasonable number of novices participated in the experiment. Novices completing this experiment is ideal, as it is desirable for the rapid scan methodology to show good results for any skill level, not just for experts.

Comments on the design and layout of the experiment were noteworthy and will be addressed in future iterations of the experiment. However, ultimately, the most important responses were the speed preferences. Of the speeds tested, 70% of respondents indicated they preferred the 2s speed with associated comments indicating 5s was boring, and 1s stressful. When asked about their preferred hypothetical speed, respondents answered most commonly with a range between 1.5s and 3s. Common requests included variable speeds to suit their preference and ability – which would be ideal outside of an experimental environment.

D. *Viability*

Given that the 2s speed was the best option of the speeds tested, it would be the ideal speed to use in a production scale flashcard analysis system.

At the 2s speed, accuracy compared to real time is 83%. Provided the requirements for rapid scan methodology are met (see the Limitations section), we argue that a 17% drop in accuracy is an acceptable trade-off for a 12× (an order of magnitude) increase in analysis speed. For Koala vocalizations in particular, they often last 20-60 seconds, fading in, reaching a climax, and then fading out. This long call means as many as 5 flashcards could have instances of the one group of vocalizations – positive identification is only necessary for one of the flashcards shown within the vocalization period.

VI. Conclusion

The experiment indicated the viability of rapidly scanning spectrograms for the basic identification of Koala vocalizations. A 12× (2.0s exposure) speedup is achievable with an acceptable trade-off in accuracy (17%).

Future work on the rapid scan methodology includes enhanced development of the interface, integration with our production website, and subsequent testing with different forms of analysis. Subsequent experimental tests could include testing: different species, different times of the day, variable exposure durations, noise-reduced spectrograms, spectrogram compression / length variation, and different numbers of classifications per flashcard.

Additionally, we think further study into the concept of a *double run analysis* of a dataset is worthwhile. By analyzing each dataset twice with a rapid scan methodology, it might be possible to decrease the drop in accuracy significantly for a trade-off of half the effective speed.

Despite the results, even when processing audio data at 12× speed, any substantial data analysis is still time consuming for a participant. We think the rapid scan methodology will be most useful when combined with multiple analysis techniques. Such techniques could include automatic filtering of the data, natural integration with our current analysis system, and some form of sampling methodology (either random or smart).

References

[1] J. Haselmayer and J. S. Quinn, "A comparison of point counts and sound recording as bird survey methods in Amazonian southeast Peru," *The Condor,* vol. 102, pp. 887-893, 2000.

[2] M. A. Acevedo and L. J. Villanueva-Rivera, "Using Automated Digital Recording Systems as Effective Tools for the Monitoring of Birds and Amphibians," *Wildlife Society Bulletin,* vol. 34, pp. 211-214, 2006.

[3] J. Wimmer, M. Towsey, B. Planitz, I. Williamson, and P. Roe, "Analysing environmental acoustic data through collaboration and automation," *Future Generation Computer Systems,* 2012.

[4] T. S. Brandes, P. Naskrecki, and H. K. Figueroa, "Using image processing to detect and classify narrow-band cricket and frog calls," *The Journal of the Acoustical Society of America,* vol. 120, p. 2950, 2006.

[5] W. Hu, N. Bulusu, C. T. Chou, S. Jha, A. Taylor, and V. N. Tran, "Design and evaluation of a hybrid sensor network for cane toad monitoring," *ACM Trans. Sen. Netw.,* vol. 5, pp. 1-28, 2009.

[6] M. A. Acevedo, C. J. Corrada-Bravo, H. Corrada-Bravo, L. J. Villanueva-Rivera, and T. M. Aide, "Automated classification of bird and amphibian calls using machine learning: A comparison of methods," *Ecological Informatics,* vol. 4, pp. 206-214, 2009.

[7] Bioacoustics Research Program. (2011, Raven Pro: Interactive Sound Analysis Software - Version 1.4 [Computer software]. Available: http://www.birds.cornell.edu/raven

[8] M. Depraetere, S. Pavoine, F. Jiguet, A. Gasc, S. Duvail, and J. Sueur, "Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland," *Ecological Indicators,* vol. In Press, Corrected Proof, 2011.

[9] J. Sueur, S. Pavoine, O. Hamerlynck, and S. Duvail, "Rapid Acoustic Survey for Biodiversity Appraisal," *PLoS ONE,* vol. 3, p. e4065, 2008.

[10] M. Towsey, B. Planitz, A. Nantes, J. Wimmer, and P. Roe, "A toolbox for animal call recognition," *Bioacoustics,* vol. 21, pp. 107-125, 2012/06/01 2012.

[11] K. A. Swiston and D. J. Mennill, "Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers," *Journal of Field Ornithology,* vol. 80, pp. 42-50, 2009.

[12] A. Taylor, G. Watson, G. Grigg, and H. McCallum, "Monitoring frog communities: an application of machine learning," 1996, pp. 1564-1569.

[13] J. Wimmer, M. Towsey, P. Roe, and I. Williamson, "Sampling environmental acoustic recordings to determine bird species richness," *Ecological Applications,* In Press.

[14] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas*, et al.*, "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey★," *Monthly Notices of the Royal Astronomical Society,* vol. 389, pp. 1179-1189, 2008.

[15] L. Sayigh, N. Quick, G. Hastie, and P. Tyack, "Repeated call types in short-finned pilot whales, Globicephala macrorhynchus," *Marine Mammal Science,* vol. 29, pp. 312-324, 2013.

[16] K.-H. Lin, X. Zhuang, C. Goudeseune, S. King, M. Hasegawa-Johnson, and T. S. Huang, "Improving faster-than-real-time human acoustic event detection by saliency-maximized audio visualization," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 2277-2280.

[17] R. J. Kosinski, "A literature review on reaction time," *Clemson University,* vol. 10, 2008.

[18] A. T. Welford, *Reaction times*: Academic Pr, 1980.

[19] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychological review,* vol. 94, pp. 115-147, 1987.

[20] S. Konishi, K. Nakajima, I. Uchida, K. Sekihara, and Y. Miyashita, "No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging," *European Journal of Neuroscience,* vol. 10, pp. 1209-1213, 1998.

[21] O. R. Joubert, G. A. Rousselet, D. Fize, and M. Fabre-Thorpe, "Processing scene context: Fast categorization and object interference," *Vision Research,* vol. 47, pp. 3286-3297, 12// 2007.

[22] W. A. Ellis, S. I. Fitzgibbon, P. Roe, F. B. Bercovitch, and R. Wilson, "Unraveling the mystery of koala vocalisations: acoustic sensor network and GPS technology reveals males bellow to serenade females," *Integrative and Comparative Biology,* vol. 50, pp. E49-E49, Jul 2010.

[23] S. FitzGibbon, W. Ellis, and F. Carrick, "Mines, farms, koalas and GPS-loggers: assessing the ecological value of riparian vegetation in central Queensland," in *10th International Congress of Ecology*, 2009.

Chapter 5
A Prototype Annotation Suggestion Tool

5.1 Introduction

The publication in this chapter details the research conducted for the decision support tool (also known as a suggestion tool) that was designed to assist the annotation process. Of the annotation processing steps (see section 3.4), the third step, classification, is the hardest step for most participants and anecdotally takes the longest to complete. This publication explores a method for aiding analysts that are classifying annotations.

For the set of data available (described further in the publication), there are some 500 types of vocalisation, which are the product of approximately 100 species. The majority of these vocalisations are generated from avian sources (many *Aves* produced more than one type of vocalisation), with some insect, marsupial, amphibian, and mammal vocalisations included.

A human analyst can discriminate between the 500 types of vocalisations; with a spectrogram and audio data as reference, an untrained participant can determine if two acoustic events are the same or not.

However, actually associating a particular vocalisation with its species name, as in being able to identify the species by memory, is a far more difficult task. In this context, the class of an acoustic event is a descriptive and unique name (either scientific or common) of the species that generated the vocalisation. Classification is easy for some well-known vocalisations like a ‘crow bark’ or a ‘kookaburra laugh’ but can be much harder for species that are not well-known.

Birding experts, biologists, and other experts can excel at recognising species by their vocalisation, using only their memory. However, the use of experts has two important limiting factors: typically, experts are experts for the species of certain areas (their knowledge is geographically constrained) and experts, by definition, are better than their peers and thus a limited resource.

It is desirable for semi-automated analysis to cater for non-experts. Allowing more analysts to participate reduces the load on other analysts and has the potential to increase overall efficiency. Anecdotal feedback from the current participants suggests amateurs are interested in participating in analysis for various reasons (general interest, benefiting their local environment).

Thus, because annotation classification is difficult for participants, users are fallible, and because it is desirable to accommodate participants with lower skill levels, it was thought necessary to create a tool that automatically assisted users’ memory (their recall ability). The decision support tool is designed to make the classification task easier for a participant by automatically suggesting annotations that are similar to an acoustic event they are currently trying to classify. This method is designed to show a shortlist of possible suggestions as analysts annotate each acoustic event, thus

reducing the recall problem-space from a memory-based 400 class problem, to a live feedback/exemplar, 5 class problem. An important goal for the suggestion tool is to provide real-time suggestions as data changes (sub-second responses) so it can be integrated directly into an annotation user interface.

The research in this publication is an initial implementation of such a system, using simple features, integrated into a user interface. This chapter directly addresses sub-research question 2 (see section 1.2): *Analysts must memorise large corpora of acoustic events to be effective; can this requirement be relaxed or reduced?*

The research for this publication required a UI prototype that was tested with quantitative and qualitative experiments. The experiment used 15 participants that demonstrated varied bioacoustic identification skills, who were contacted both directly and via email. The research group's ethics policy (detailed in section 1.6) applies to this chapter. Both expert and amateur participants were used. Participants used the interface and their performance was measured. Qualitative feedback was collected through a short paper-based survey. A UI artefact was produced as part of the research.

Results for the initial study suggested participants liked the idea of a suggestion tool but found the implementation and performance inadequate. The decision support tool has a few basic limitations. The tool relies on previous data from participants. It cannot suggest correct class until at least one example vocalisation has been annotated. However, this problem can be mitigated by ensuring experts conduct the initial analysis on new datasets.

5.2 Conference Paper – Large Scale Participatory Acoustic Sensor Data Analysis: Tools and Reputation Models to Enhance Effectiveness

Truskinger, A., Yang, H. F., Wimmer, J., Zhang, J., Williamson, I., & Roe, P. (2011). *Large Scale Participatory Acoustic Sensor Data Analysis: Tools and Reputation Models to Enhance Effectiveness*. Paper presented at the 2011 IEEE 7th International Conference on E-Science (e-Science), Stockholm. http://dx.doi.org/10.1109/eScience.2011.29

This conference paper has been peer reviewed and published. This paper was primarily produced by two authors: **Anthony Truskinger** and Hao-fan Yang. Writing a thesis by publication requires that the papers included in the thesis be done so verbatim. The sections in this paper regarding suggestion tools are the research of **Anthony Truskinger**. The sections of the paper regarding trust and reputation models are the work of Hao-fan Yang.

5.3 Statement of Contribution

RESEARCH STUDENTS CENTRE
Examination Enquiries: 07 3138 1839
Email: research.examination@qut.edu.au

Statement of Contribution of Co-Authors for Thesis by Published Paper

The authors listed below have certified* that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter:

Publication title and date of publication or status:

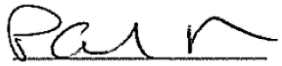
Large Scale Participatory Acoustic Sensor Data Analysis:
Tools and Reputation Models to Enhance Effectiveness

Published 2011

Contributor	Statement of contribution*
Anthony Truskinger	Wrote the manuscript, designed the experiment, created the software prototype
Signature 	
Date 21/05/2015	
Haofan Yang*	Helped write the paper, helped to design the experiment, designed the trust model
Jason Wimmer*	Helped to write the paper
Jinglan Zhang*	Supervisor – Oversaw and contributed to the entire paper
Ian Williamson*	Supervisor – Oversaw and contributed to the entire paper
Paul Roe*	Supervisor – Oversaw and contributed to the entire paper

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Paul Roe		20/5/15
Name	Signature	Date

Chapter 6
Decision Support for the Efficient Annotation of Bioacoustic Events

6.1 Introduction

The publication in this chapter details the additional research conducted for the decision support tool (also known as the suggestion tool) used for supporting annotating analysts. The results of Chapter 5 show that the rate at which participants annotated new events increased and that the participants liked the concept of the suggestion tool. However there were two distinct limitations in the original prototype: first, the suggestion performance (accuracy) of the tool was not sufficient. Accuracy of the tool was not directly measured by the experimental methodology; it instead measured the performance of the participants, who then commented on the low accuracy of the tool. Second, experiment participants remarked in the survey that they found the tool awkward to use – it was not sufficiently integrated into the annotation UI.

Thus, the three goals of the additional research in this publication were to:

- measure baseline suggestion performance of the tool in the original publication
- increase the suggestion performance significantly, and
- ensure the tool can remain responsive after improvements.

Additionally, an investigation of a better-integrated version of the decision support tool was conducted. This chapter (along with Chapter 5) directly addresses sub-research question 2 (see section 1.2): *Analysts must memorise large corpora of acoustic events to be effective; can this requirement be relaxed or reduced?*

The research was conducted as an exploratory analysis of varying algorithmic techniques that would improve the performance of the decision support tool. All reported results for this research are quantitative and did not utilise participants. The research group's ethics policy (detailed in section 1.6) applies to this chapter.

The performance of the suggestion tool was evaluated based on sensitivity results for test data on which the suggestion tool was applied. A dataset of 82 000 annotations was exported from the database and split into test and training sets. All experiments were conducted automatically by simulation and all were deterministic (except for the randomised control cases). Results for the exploratory analysis rigorously demonstrated a doubling in the suggestion tool's performance whilst maintaining acceptable response times.

Additional data from the experiments, not included in the publication, are included in *Appendix D – Additional Suggestion Tool Results*.

6.2 Journal Paper – Decision Support for the Efficient Annotation of Bioacoustic Events

Truskinger, A., Towsey, M., & Roe, P. (2015). Decision Support for the Efficient Annotation of Bioacoustic Events. *Ecological Informatics, 25*, 14-21. doi: 10.1016/j.ecoinf.2014.10.001

This journal article has been peer reviewed and published in *Ecological Informatics* journal.

6.3 Statement of Contribution

RESEARCH STUDENTS CENTRE
Examination Enquiries: 07 3138 1839
Email: research.examination@qut.edu.au

Statement of Contribution of Co-Authors for Thesis by Published Paper

The authors listed below have certified* that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter:

Publication title and date of publication or status:

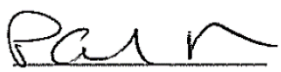
Decision Support for the Efficient Annotation of Bioacoustic Events

Published, 2015

Contributor	Statement of contribution*
Anthony Truskinger	Wrote the manuscript, created the software prototype, conducted experiments
Signature 	
Date 21/05/2015	
Michael Towsey*	Cowriter – Edited the written parts of the research and contributed to the theory.
Paul Roe*	Supervisor – Oversaw and contributed to the entire paper

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Paul Roe  20/5/15

Name Signature Date

Chapter 7
Tag Cleaning and Linking

7.1 Introduction

The publication in this chapter details the research conducted for the cleaning and linking of a set of corrupted tags. Annotations (which have tags) are the main output of the analyses applied to acoustic data. The annotation data is sent to ecologists, who in turn use the data to answer ecological questions; it is important that the data sent is consistent, rigorous, and ultimately usable.

There are three main types of error associated with the data output by annotations: inconsistent segmentation; incorrect classification (or incorrect event/tag association); and textually incorrect tags. An annotation is incorrectly segmented, if the bounds of the annotation do not include the entire acoustic event that is being annotated. An annotation is incorrectly classified when, for example, a Torresian crow (*Corvus orru*) acoustic source may be labelled (associated) with the "laughing kookaburra" *common name* tag. Finally, an annotation could be considered incorrect if the tags associated with it are textually incorrect, for example, spelt incorrectly. These error classes are not mutually exclusive.

The research in this chapter is focussed on tags and automatically fixing the various textual problems that can occur in a folksonomic tagging system. This research addresses the third stage of annotation (classification (see section 3.4)) and the third sub research question (see section 1.2): *Can human generated folksonomies used to tag acoustic events be mapped back to taxonomies?*

Correcting tags is important because many of the analyses conducted on the annotation data rely on summarising the frequencies (occurrence counts) of different tags. Typically, annotations share a relatively small set of tags. For example, of the 60 746 annotations in the training dataset used in Chapter 6, there are only 382 unique tags. This means a few malformed tags are: a) difficult to find and correct within a large set and b) can have a significant effect on the groups formed when summarising the annotation data.

Choosing a folksonomy for tagging acoustic events (to form annotations) was a conscious choice for the host project. However, for their stated advantages, folksonomies also have disadvantages. After much annotation was done, it was determined that, ideally, the support of a hybrid folks-taxonomy would be a better alternative.

The research in this chapter has three contributions:

- A method for correcting corrupted tags was developed.
- That method was then used to link folksonomic tags to formal taxa.
- A widget was designed to take advantage of the newly cleaned data.

The maintenance of the tags has allowed for the transformation of the entire annotation dataset resulting in a dataset that takes far less manual effort by ecologists users to clean. After the publication of these results, the cleaning process was applied to the QUT Ecoacoustics' production database – it cleaned and replaced all tags in that dataset (130K annotations).

The research in this chapter conducted a post-hoc analysis of data generated by participatory analysis. The research is exploratory (posteriori) in nature with performance measured quantitatively. An algorithm was designed to check for and correct problems in tags using various techniques. Summary statistics were used to highlight the resulting changes in the dataset. A dataset of 90 255 annotations was exported from the database. The rules for correcting the tags associated with annotations were deterministic. The resulting rules created a series of software artefacts, heuristics, that can be applied to tags in the future to prevent further corruption. No participants were needed for this research as the research was conducted with data only. With permission, data was extracted from the QUT Ecoacoustics website. No identifying data was needed or exported for this work. The research group's ethics policy (detailed in section 1.6) applies to this chapter.

Of the 90 225 annotations, 87% of their tags were cleaned/repaired in some way. Additionally, 85% of the dataset was associated with a formal species name allowing for the linking and retrieval of external data for those annotations. As artefacts, an information widget and a set of heuristics were created. Additional data, scripts, code, or results can be obtained by contacting the author.

7.2 Conference Paper – Reconciling Folksonomic Tagging with Taxa for Bioacoustic Annotations

Truskinger, A., Newmarch, I., Cottman-Fields, M., Wimmer, J., Towsey, M., Zhang, J., & Roe, P. (2013). *Reconciling Folksonomic Tagging with Taxa for Bioacoustic Annotations*. Paper presented at the 14th International Conference on Web Information System Engineering (WISE 2013), Nanjing, China. http://dx.doi.org/10.1007/978-3-642-41230-1_25

This conference paper has been peer reviewed and published.

7.3 Statement of Contribution

RESEARCH STUDENTS CENTRE
Examination Enquiries: 07 3138 1839
Email: research.examination@qut.edu.au

Statement of Contribution of Co-Authors for Thesis by Published Paper

The authors listed below have certified* that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

In the case of this chapter:

Publication title and date of publication or status:

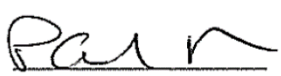
Reconciling Folksonomic Tagging with Taxa for Bioacoustic Annotations

Published, 2013

Contributor	Statement of contribution*
Anthony Truskinger Signature  Date 21/05/2015	Wrote the manuscript, designed the experiment, guided the software implementation
Ian Newmarch*	Created the software
Mark Cottman-Fields*	Helped to write the manuscript
Jason Wimmer*	Helped to write the manuscript
Michael Towsey *	Supervisor – Oversaw and contributed to the entire paper
Jinglan Zhang *	Supervisor – Oversaw and contributed to the entire paper
Paul Roe*	Supervisor – Oversaw and contributed to the entire paper

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Name	Signature	Date
Paul Roe		20/5/15

Chapter 8
Conclusions

This thesis posed the question "*How can automation improve the efficiency of manual analysis of faunal acoustic events in recorded acoustic data?*". To address this question, this thesis contributes a set of efficiency improving techniques for use in a semi-automated faunal annotation system for acoustic sensor data. These contributions have been published as a series of papers that, as per the format of thesis by publication, have been included verbatim in this thesis. The publications themselves each contribute to knowledge independently, yet when considered as a whole, produce a cohesive result. Each paper formed a major chapter in this thesis and was prefixed by an introduction that provided the necessary context to understand the publication's place within the thesis.

This conclusion summarises the motivations, research questions, and methodology used for this research. Importantly, the publications' contributions are summarised and presented as a single cohesive contribution to knowledge.

8.1 Motivations

Monitoring the environment is an important part of understanding the world we live in. Of the various environmental monitoring methods available to scientists, this research focussed on terrestrial acoustic monitoring of the environment via acoustic sensors. Acoustic sensors allow researchers to monitor the environment over large spatiotemporal scales. The data collected is a permanent, unbiased, record for the area where data was captured.

This thesis was motivated by the need to analyse acoustic sensor data for faunal vocalisations. The results of analysis can be provided to ecologists so that they can make ecological inferences –faunal vocalisations are used as proxies for other biodiversity metrics.

As seen in the literature, monitoring the environment with sensors is a common activity. In particular, acoustic sensors are used to monitor *Aves*, *Chiroptera*, and to a lesser extent *Anurans*. When considering the analysis of faunal events in audio data, all literature presented falls somewhere in the spectrum of automated to manual analysis. Fully automated, high-accuracy, solutions for the analysis of acoustic sensor data are an ideal solution but are currently considered an intractable problem.

While automated analysis continues to improve, in the interim, there is value in analysing data manually. Fully manual analysis requires inordinate amounts of human time and effort. Despite this, the data obtained is valuable: it can be used by ecologists to address smaller scale research questions and be used to enhance the development of automated approaches. Because human analyst resources are limited, it is important to use them efficiently. In particular, the need for

analysts with domain-relevant skills limits the pool of participants that can contribute. While human analysts continue to be needed, it was hypothesised that the methods of analysis with the most potential would combine human and computation analysis – those that take advantage of the complementary skills available to each. This combination of computational and human processing is termed *semi-automated analysis*.

This thesis used resources from the QUT Ecoacoustics Research group. The bioacoustics software package the research group produces, the *bioacoustic workbench*, is a distributed web-based framework that allows playback, visualisation, and annotation of acoustic data to be conducted digitally. When reviewing the bioacoustic workbench, it was observed that too much of the work that human participants did was mundane, unnecessary, or better suited for a machine. The participants, software, and data resources of the QUT Ecoacoustics Research group were used to host the experiments of this thesis.

8.2 Research Questions

The core research question for this thesis is:

How can automation improve the efficiency of manual analysis of faunal acoustic events in recorded acoustic data?

In this context, the analysis of acoustic faunal events produces annotations. There are three steps required to create an annotation (a bounded, labelled, acoustic event):

1. Detection: in voluminous sensor data, there are acoustic events of interest that must be first found before they can be processed.
2. Segmentation: the bounds of the event must be defined, so that signal that is part of the event is clearly marked in time and frequency domains.
3. Classification: actually deciding what produced the acoustic event of interest. The classification is applied as a set of tags to the annotation.

For the aforementioned annotation process, this thesis has investigated three methods for enhancing the efficiency of participants in semi-automated faunal acoustic event annotation systems. These methods map to the sub-questions identified in section 1.2:

1. Can the faunal event detection speed of analysts be enhanced?
2. Analysts must memorise large corpora of acoustic events to be effective; can this requirement be relaxed or reduced?
3. Can human generated folksonomies used to tag acoustic events be mapped back to taxonomies?

The mapping between annotation steps, sub research questions, and chapters can be seen in Table 3.

Table 3 – The mapping between annotation steps, sub research questions, and thesis chapters

Annotation Step		**Sub Research Question**	**Chapter (s)**
Detection		1	4
Segmentation		N/A	
Classification	Sub step: class recall	2	5 & 6
	Sub-step: labelling	3	7

For annotation step 1, detection, Chapter 4, Rapid Scanning of Spectrograms enhanced the detection speed of users and addressed sub-research-question 1.

For annotation step 2 of the annotation process (segmentation), defining the bounds of an acoustic event was found to be easy for humans. Human participants can discern the time and frequency bounds of an acoustic event and draw those bounds around the event (on a spectrogram) within a few seconds; humans can do this for noisy audio data mixed with overlapping signals. The literature shows that currently, humans perform this task far better than their machine equivalents. Consequently, no research in thesis focussed on improving the already sufficient efficiency of humans at defining the dimensions of an acoustic event.

For annotation step 3, the literature shows that classifying an acoustic event is difficult for both machines and humans. Classification of a faunal acoustic event is a two-step process: class recall and applying a class label. Typically, these steps are not distinguished – in a supervised machine learning process; training data is associated with a class label, making the distinction trivial. This thesis drew the distinction based on the skills of human analysts. For the class recall stage of annotation classification, Chapters 5 and 6 used existing annotations to create a decision support tool to assist users. For the labelling stage of annotation classification, Chapter 7 researched methods of cleaning and keeping clean the tag folksonomies used in annotation.

8.3 Findings

8.3.1 Rapid Scanning of Spectrograms (Event Detection)

The rapid scan methodology stemmed from the observation of analysts employing a similar behaviour to rapidly find acoustic events of interest. They would drag the navigation seek bar rapidly in the forward direction which in turn ‘fast-forwarded’ the spectrogram animation displayed to

them. During this process, the analysts could not hear audio but could still visually discern events of interest.

Given that this emergent behaviour was theoretically sound, it was formally tested to see if the technique was useful. The goal of the experiment was to see how effective human analysts were at filtering out irrelevant sections of audio, based on quick exposure to spectrograms. A suitable interface was designed specifically to assist the participants in quickly identifying acoustic events.

The prototype UI was developed within an experimental test framework. Instead of animating the spectrogram in a sliding fashion, spectrograms were shown as fixed, static, individual images. These images were displayed for fixed intervals of time and the participant was required to respond with either a confirmation (the target pattern was identified) or rejection (the target pattern was not identified) for each slide.

The experiment measured the spectrograms' exposure time versus participant accuracy. Each participant was treated to three exposure times: 5, 2, and 1 second exposures, with effective rapid scanning speeds of 4.8×, 12×, and 24× respectively (24 second spectrograms). To make the experiment fair and consistent it only tested for the detection of one vocalising species, the male koala (*Phascolarctos cinereus*).

The analysis of the data revealed that the 2 second exposure speed was the most effective speed and also the speed most liked by participants – not too boring, not too quick. Although the 2 second speed reported a drop in accuracy, it was within tolerable limits. The results were subjected to *repeated measures ANOVAs* to ensure statistical relevancy – the drop in performance between the 2 second and 1 second speeds was statistically significant.

The results of the rapid scanning paper are promising – participants' time efficiency in identifying acoustic events was increased by a factor of 12 at a two-second exposure. The results indicate that such a method could be used in a real annotation system to help participants quickly identify acoustic events of interest, whilst filtering out irrelevant data.

However, more work is needed to assess the generalisability of the rapid scan methodology. The experiment showed that the method works with one type of vocalisation – a koala bellow. More experimentation is needed to see if other suitable vocalisations will experience the same detection rate. Another, similar, limitation in this experiment is that only one type of acoustic event was searched for by analysts. In the general case, this tool would be more useful for detecting more than just one type of acoustic event in the source data – the ideal question posed to analysts would be,

"Are there any interesting acoustic events in this image?". Anecdotally, this task is expected to be easier for participants; but again, more experimentation is needed.

8.3.2 Decision Support Tool (Class Recall)

Measuring similarity determines how similar a target unknown event is to a knowledge base of known events; for a human analyst this is memory, for a machine it is training data. The literature demonstrated that humans are good at determining similarity between two instances, due to their ability to use creative and qualitative features.

However, trying to remember which class (species) an unknown acoustic event belongs to is more difficult for a human participant. Citizen experts are participants that have excellent recall of faunal acoustic event types, gained from years of experience; they are proficient at the task of associating vocalisations to species names. Yet requiring experts – a limited resource – for analysis is not necessary for the rest of the annotation process. This problem of recalling what species vocalised, of hundreds of types of vocalisation, was addressed in this thesis with a decision support tool designed to allow non-experts to classify acoustic events.

8.3.2.1 Chapter 5

It was theorised that a suggestion tool, in the spirit of an auto-complete box, would help improve the efficiency of participants, especially those not familiar with all of the types of vocalisations from a region.

An initial implementation of a decision support tool used a simple algorithm and a small, high quality, training dataset in an experiment. Users had to draw an annotation, and then click the suggestion button to get results that were displayed in a separate window. The experiment saw no significant change in participant accuracy or time taken. However, there was a significant improvement in the number of annotations created by users – particularly for novice users.

The qualitative portion of the experiment revealed users thought the idea of the decision support tool had promise but they thought the tool needed to be improved. Participants stated that the tool's accuracy was not sufficient and that the user interface was awkward.

The experiment did not evaluate the accuracy of the decision support tool but rather, the performance of the participants using it. Post-paper analysis revealed that the poor performance of the analysis tool was a problem; for the top five suggestions shown to a user, there was only a 25% chance that the correct suggestion would be shown. On reflection, this result was unsurprising; using a small training dataset with un-normalised features in a machine learning style problem should not be expected to perform well.

8.3.2.2 Chapter 6

The research that followed addressed the identified shortcomings of the original decision support tool. The aim of this new research was to increase the suggestion performance (the accuracy) of the decision support tool and to incorporate the tool into a better user interface.

Typically, performance is enhanced by adding more training data. The training dataset was increased to 60 000 ordinary annotations up from the 400 high quality annotations previously used. Using the simplest algorithm (Euclidean similarity search with three *bounding box* features) increased performance substantially: sensitivity for returning a correct suggestion within five suggestions increased to 64% from 25%. However, the computation time needed for this simple algorithm was excessive and rendered the usage of the tool useless in an interactive scenario. Computational performance was profiled as $O(n)$ – scaling linearly with the number of training data instances added.

Research was continued to create a better algorithm or set of features, that not only provided better suggestions but that also scaled with large amounts of training data. A series of potential algorithms and features were considered, followed by the setup of a test-protocol for their combinations. Hundreds of combinations were tested; the best result (a trade-off between accuracy and computational performance) was the Euclidean Distance similarity metric, matching test annotations to Z-Score normalised, class prototypes, using the three dimensional features of an annotation (start frequency, end frequency, and duration), while not making use of the 'time of day' feature.

The new algorithm and feature set demonstrated an acceptable increase in suggestion performance (48.12% compared to 24.56% for the top five suggestions). It did not perform as well for suggestions when compared to the basic algorithm. However, in terms of computational performance, the improved algorithm was two orders of magnitude better than the basic algorithm; it returned five suggestions in 55ms compared to the 3.2s of the slower algorithm. Importantly, the improved algorithm scaled logarithmically ($O(log\ n)$) with training data.

The result is a sub-(deci)second decision support tool that has 48% chance (from five suggestions) of suggesting the correct class of acoustic event. The research also embedded the decision support tool into a prototype interface that automatically provided suggestions when a user started annotating an acoustic event.

The goals of the decision support research were reached, yet, there are additional research questions to be investigated. Ideally, the suggestion performance of the algorithm would be better: additional training data and better algorithmic techniques are expected to enhance performance.

Including more features is a logical extension to this project. Given datasets from larger spatiotemporal distributions, contextual features should prove useful in discriminating interesting events. Additional potential for this technique lies in extracting features from the spectrogram and audio signal of the sections of audio bounded by the annotation.

Additionally, the results of the analysis are sensitive to how users annotate – particularly the heuristics of individuals that govern their drawing of bounding boxes. Further study is needed on the effect of inter-user variance. Lastly, applying this tool to datasets from different ecosystems is necessary to evaluate the generalisability of the decision support technique for faunal acoustic event monitoring.

8.3.3 Tag Cleaning and Linking (Labelling)

The second stage to classification (and last step of annotating) is the application of a class label to the annotation; the task of literally applying a textual label can produce a surprising number of errors. The user knows what they are trying to label; they just have to type the tag correctly. In a system that allows free form tagging (a folksonomic approach) there are more opportunities to make basic textual mistakes. The literature showed that these problems with tagging systems are common.

On reviewing the annotation data used by this thesis, it was observed that textual errors were prevalent, resulting in inconsistent and sometimes incorrect data. When this data is exported to ecologists, the result is repetitive, inefficient cleaning undertaken by them. To solve these inefficiencies, research was conducted to find a method of cleaning and keeping clean the folksonomic tags.

The research that followed produced a method for repairing and reconciling a damaged faunal tag folksonomy through the use of a formal species taxonomy. The cleaning and repairing of the tag set (the folksonomy) was necessary but linking the folksonomy, particularly the common and species name tags, to a taxonomy, represented an additional opportunity to then utilise external data sources.

The cleaning and linking algorithms (a combination of heuristics and spell checking algorithms) were applied to a 90 225 instance tag dataset. Normalisation was required for 87% of the tags and more advanced error correction was required for 1.12% of the tags. The result of 95% of the common name tags being associated with species names was a successful, automated cleaning of the tag data.

To demonstrate the usefulness of linking the folksonomic tags to a taxonomic data source, a UI widget prototype was developed. This widget used the cleaned and linked tag data to retrieve, in real time, additional information about the tag that was typed. For species, structured data that included statistics like geographical distribution, seasonal variation, migrations patterns, and even images, were returned to assist an annotating participant.

The widget that was created and the cleaned tag data were incorporated into the QUT Ecoacoustics research group's database. In particular, the same heuristics used to clean the tag data, were also applied as validation heuristics of the folksonomic tags, as they were applied to annotations – thus decreasing the possible errors that could be made by a human in future annotation tasks. The tag cleaning and linking research is a specific solution to a problem the QUT Ecoacoustics research group had; this means it has limited applicability to other fields of study. These cleaning techniques may be applicable to other datasets where a folksonomy was initially used for usability, but where an effective taxonomy exists already.

8.4 Conclusion

The aforementioned findings were published as individual works. With the description of the annotation steps (detection, segmentation, and classification) it was shown that these contributions were part of the larger semi-automated annotation narrative.

8.4.1 Relevance to Literature and Implications

The literature review revealed five conclusions: acoustic sensor recordings are used to monitor the environment; identifying fauna within those recordings can form ecological conclusions; automated methods for doing this are intensely researched but not yet capable of providing a complete solution; and finally, humans have excellent classification skills but these need to be used efficiently.

Chapters 4, 5, and 6 demonstrate how humans can be assisted with automation. The core concept of these ideas is to reduce strain (monotonous work) on users and instead only involve a user when classification is needed. As an important side effect, automated assistance (particularly for class recall activities) should lower the skill threshold required for human analysts; examples of this effect were measured in Chapters 4 and 5.

The research in Chapter 7 demonstrates the work required to reconstitute corrupt data when appropriate computational support was not provided to users. The corrupted and inconsistent folksonomy could have been avoided if appropriate verification and well-defined tagging practices were used originally. However, the research done to restore integrity to the tag set provided new opportunities to explore and link a folksonomy to external data sources.

In summary, the contribution to knowledge provided by this thesis is that **automation can improve the efficiency of manual analysis of faunal acoustic events**. The implications of this new knowledge mean that other eScience projects that rely on data collection techniques may be able to reuse this philosophy: a semi-automated system can produce valuable, effective data – if its users are appropriately and efficiently supported. This literature shows that this observation has been demonstrated by other projects, particularly those fostered by the Zooniverse organisation. However, this is the first time a study of semi-automated analysis of this magnitude has been focussed on terrestrial faunal acoustic event identification from sensor data.

As a secondary implication, it is suggested there is utility in incorporating participants into automated methods of analysis. The literature cited examples of organisations that questioned the utility of incorporating human analysts into traditionally automated methods of analysis. Typically the contention centres on scaling analysis: fully automated analysis can be scaled with just compute resources, whereas semi-automated analysis is bounded by the number and quality of human analysts available. The efficiency of the human analysts – the subject of this thesis – determines how much data they output.

Even though human analysts produce a fraction of the data of machines, the data analysed is still valuable. For example, small amounts of analysed data can still address smaller scale ecological questions. However, based off the findings in this thesis, it is suggested that the real value in semi-automated methodologies is in using human analysts to assist automated methods. This *feedback loop* concept, where assistance is alternately applied to both parts of semi-automated methodologies, has potential. Automated methods, especially those utilising machine learning techniques, generally perform better with more training data. The techniques described in this thesis, particularly the decision support tool and rapid scanning spectrograms, allow datasets to be labelled more efficiently than their manual equivalents. Human analysts can be used to bootstrap new data sources, analyse training data, and reinforce learning algorithms through class disambiguation. Additionally, after analysis, semi-automated methods have the potential to verify results.

8.4.2 Limitations and Further Research

It is important to reflect on the limitations of the research done. Each of the chapters listed their own limitations, however for the thesis overall, there is one significant limitation: integrating each of the studied techniques together into one system, to test the overall effect on analyst efficiency, was not completed. The primary reason for this limitation is the amount of programming (non-research) work that was required. The host software used to test this thesis's concepts is production grade

software. Small experiments and prototypes can be attached to the software in isolation and tested but then must be removed. Testing all components requires them to be properly integrated and developed for reliability – work that is outside the scope of this thesis. The assertion is made that the main research question was sated; that the efficiency of users was improved is thus necessarily asserted transitively: through a set of smaller efficiency gains, it can be assumed that the overall system has improved in efficiency. However, there exists a possibility that combining all these techniques together may not actually result in an overall improved efficiency. If these techniques are applied in concert, they should ideally be applied individually and incrementally, with careful measurement – just like any other experiment.

The second major limitation to this thesis was the limited scope of available data. The methodologies tested in this thesis depend on large datasets of acoustic sensor data and annotations. Datasets with these properties are not common. The data obtained for use in this thesis had two significant limitations: it was generated by a small community of human analysts and the majority of the audio data that was analysed was from one geographical location. Future work should include studies on inter and intra user variance for the different types of analysis tasks available to human analysts of bioacoustic data.

Additionally, determining the applicability of the techniques in this thesis, for acoustic sensors data (and the fauna within) for other regions is important aspect of assessing generalisability. The rapid scanning methodology is expected to remain useful in different regions, provided the questions asked of analysts is appropriate – an appropriate question is on that tasks analysts to detect large scale, macro detail in spectrogram images (as opposed to minutia). For the decision support tool, it is expected that it will again be useful when applied to other ecosystems. However, the effectiveness of the tool is dependent on the temporal and frequency distributions of the bioacoustic events of the fauna present in the ecosystem – the more varied the bioacoustic events, the better the decision support tool will work.

Each of the chapters provides avenues of further research that can be pursued. Generally, though, for the thesis as a whole, there are two important questions to consider: are the techniques presented within applicable to (1) other bioacoustics software packages; and (2) to other types of eScience data analysis problems?

Other bioacoustic software packages can reuse the methodologies presented in this thesis. In particular, the *Pumilio* project shares similar goals to that of the *QUT Ecoacoustics research group* – all techniques presented are compatible with their software, however, they would need to embrace the concept of semi-automated analysis. Additionally, *xeno canto*, could benefit from the decision

support tool for helping their users classify an unknown event. However, xeno canto will not benefit from the rapid scanning methodology or tag cleaning; the recordings uploaded to xeno-canto are short (no need for rapid scanning) and they use a formal taxonomy to classify acoustic events. It remains unclear how other projects, like *Raven* or *Songscope*, can benefit from the techniques presented in this thesis.

In summary, based on the literature surveyed and the results from this thesis's experiments, this thesis recommends that in general terms, wherever possible, human analysts using computers should be assisted. Assisted users make less mistakes and are more efficient. Computer assistance is particularly useful in high-class classification problems – like that of terrestrial bioacoustic event identification. The classification of bioacoustic events in acoustic sensor data exhibits properties that are amenable to semi-automated assistance.

Bibliography

Abdulmonem, A., & Hunter, J. (2010). *Enhancing the Quality and Trust of Citizen Science Data*. Paper presented at the IEEE Sixth International Conference on e-Science, Brisbane. http://doi.ieeecomputersociety.org/10.1109/eScience.2010.33

Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., & Aide, T. M. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics, 4*(4), 206-214. doi: 10.1016/j.ecoinf.2009.06.005

Agranat, I. (2009). *Automatically Identifying Animal Species from their Vocalizations*. Paper presented at the Fifth International Conference on Bio-Acoustics, Holywell Park. http://bioacoustics2009.lboro.ac.uk/abstract.php?viewabstract=57

Agranat, I. (2013). *Bat species identification from zero crossing and full spectrum echolocation calls using Hidden Markov Models, Fisher scores, unsupervised clustering and balanced winnow pairwise classifiers.* Paper presented at the Proceedings of Meetings on Acoustics.

Aide, T. M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., & Alvarez, R. (2013). Real-time bioacoustics monitoring and automated species identification. *PeerJ, 1*, e103. doi: 10.7717/peerj.103

Allman, J. M. (2000). *Evolving brains*. New York: Scientific American Library.

Alpaydin, E. (2004). *Introduction to machine learning*: MIT press.

Audacity Team. (2013). Audacity 2.0.3 (Version 2.0.3). Retrieved from http://audacity.sourceforge.net/

Bagwell, C. (2013). SoX-Sound eXchange (Version 14.4.1). Retrieved from http://sox.sourceforge.net/

Bardeli, R. (2009). Similarity Search in Animal Sound Databases. *IEEE Transactions on Multimedia, 11*(1), 68-76. doi: 10.1109/TMM.2008.2008920

Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K. H., & Frommolt, K. H. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters, 31*(12), 1524-1534. doi: 10.1016/j.patrec.2009.09.014

Bioacoustics Research Program. (2011). Raven Pro: Interactive Sound Analysis Software - Version 1.4 [Computer software]. http://www.birds.cornell.edu/raven

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication*.

Brandes, S. (2008). Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International, 18*(Supplement S1), S163-S173, M163 - 110.1017/S0959270908000415.

Brandes, T. S., Naskrecki, P., & Figueroa, H. K. (2006). Using image processing to detect and classify narrow-band cricket and frog calls. *The Journal of the Acoustical Society of America, 120*, 2950.

Bridle, J., & Brown, M. (1974). An experimental automatic word recognition system. *JSRU Report, 1003*, 5.

Brown, M., Chaston, D., Cooney, A., Maddali, D., & Price, T. (2009). *Recognising birds songs-comparative study*. Unpublished manuscript, University of Sheffield. Retrieved from https://wiki.dcs.shef.ac.uk/wiki/pub/Darwin2009/WebHome/jasa.pdf

Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., & Srivastava, M. B. (2006). Participatory Sensing. *In ACM Sensys workshop on WorldSensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications*, 117-134. doi: 10.1.1.122.3024

Butler, R., Servilla, M., Gage, S., Basney, J., Welch, V., Baker, B., . . . Freemon, D. (2007). Cyberinfrastructure for the analysis of ecological acoustic sensor data: a use case study in grid deployment. *Cluster Computing, 10*(3), 301-310.

Catchpole, C., & Slater, P. (2008). *Bird song: Biological themes and variations* (2nd ed.). Cambridge: Press Syndicate University of Cambridge.

Chávez, E., Navarro, G., Baeza-Yates, R., & Marroquín, J. L. (2001). Searching in metric spaces. *ACM Comput. Surv., 33*(3), 273-321. doi: 10.1145/502807.502808

Chesmore, D. (2007). 6 The Automated Identification of Taxa: Concepts and Applications. *Automated Taxon Identification in Systematics: Theory, Approaches and Applications*, 83.

Chesmore, E. D., & Ohya, E. (2004). Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition. *Bulletin of Entomological Research, 94*(04), 319-330.

Christidis, L., Boles, W., & Ornithologists' Union, R. A. (1994). *The taxonomy and species of birds of Australia and its territories*: Royal Australasian Ornithologists Union.

Clements, J. (2007). *The Clements checklist of birds of the world*: Comstock Pub. Associates/Cornell University Press.

Clements, J., Schulenberg, T., Iliff, M., Sullivan, B., Wood, C., & Roberson, D. (2012). The eBird/Clements checklist of birds of the world: Version 6.7 (Version 6.8). Retrieved from http://www.birds.cornell.edu/clementschecklist/download/

Cohn, J. P. (2008). Citizen Science: Can Volunteers Do Real Research? *Bioscience, 58*(3), 192-197. doi: 10.1641/bs80303

Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., Shirk, J., & Bonney, R. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *Bioscience, 59*(11), 977-984. doi: 10.1525/bio.2009.59.11.9

Cottman-Fields, M., Truskinger, A., Wimmer, J., & Roe, P. (2011). *The Adaptive Collection and Analysis of Distributed Multimedia Sensor Data*. Paper presented at the 2011 IEEE 7th International Conference on E-Science (e-Science), Stockholm.

Cuff, D., Hansen, M., & Kang, J. (2008). Urban Sensing: Out of the Woods. *Communication of the ACM, 51*(3), 24-33.

Cugler, D. C., Medeiros, C. B., & Toledo, L. F. (2011). *Managing animal sounds-some challenges and research directions.* Paper presented at the Proceedings V eScience Workshop-XXXI Brazilian Computer Society Conference.

Culverhouse, P. F., Williams, R., Reguera, B., Herry, V., & González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine indentification of dinoflagellates. *Marine Ecology Progress Series, 247*, 17-25. doi: 10.3354/meps247017

Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., & Sueur, J. (2012). Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland. *Ecological Indicators, 13*(1), 8. doi: http://dx.doi.org/10.1016/j.ecolind.2011.05.006

DIN ISO 9613-1. (1993). 9613–1: 1993. Acoustics. Attenuation of sound during propagation outdoors. Part 1: Calculation of the absorption of sound by the atmosphere *International Organization for Standardization, Geneva*.

Dong, X., Towsey, M., Jinglan, Z., Banks, J., & Roe, P. (2013). *A Novel Representation of Bioacoustic Events for Content-Based Search in Field Audio Data*. Paper presented at the 2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Hobart.

Doupe, A. J., & Kuhl, P. K. (1999). BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms. *Annual Review of Neuroscience, 22*(1), 567-631. doi: doi:10.1146/annurev.neuro.22.1.567

Duan, S., Towsey, M., Zhang, J., Truskinger, A., Wimmer, J., & Roe, P. (2011). *Acoustic component detection for automatic species recognition in environmental monitoring.* Paper presented at the 2011 Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Adelaide.

Duan, S., Zhang, J., Roe, P., Towsey, M., & Buckingham, L. (2012). *Timed and probabilistic automata for automatic animal Call Recognition.* Paper presented at the 2012 21st International Conference on Pattern Recognition (ICPR), Tsukuba.

Duan, S., Zhang, J., Roe, P., Wimmer, J., Dong, X., Truskinger, A., & Towsey, M. (2013). *Timed Probabilistic Automaton: A Bridge between Raven and Song Scope for Automatic Species Recognition.* Paper presented at the Twenty-Fifth IAAI Conference, Bellevue.

Echarte, F., Astrain, J., Córdoba, A., & Villadangos, J. (2008). Pattern Matching Techniques to Identify Syntactic Variations of Tags in Folksonomies. In M. Lytras, J. Carroll, E. Damiani & R. Tennyson (Eds.), *Emerging Technologies and Information Systems for the Knowledge Society* (Vol. 5288, pp. 557-564): Springer Berlin Heidelberg.

Ellis, W. A., Fitzgibbon, S. I., Roe, P., Bercovitch, F. B., & Wilson, R. (2010). Unraveling the mystery of koala vocalisations: acoustic sensor network and GPS technology reveals males bellow to serenade females. *Integrative and Comparative Biology, 50*, E49-E49.

Feyyad, U. M. (1996). Data mining and knowledge discovery: making sense out of data. *IEEE Expert, 11*(5), 20-25. doi: 10.1109/64.539013

Frommolt, K., Tauchert, K., & Koch, M. (2008, December 2007). *Advantages and Disadvantages of Acoustic Monitoring of Birds—Realistic Scenarios for Automated Bioacoustic Monitoring in a Densely Populated Region, Computational bioacoustics for assessing biodiversity.* Paper presented at the Proceedings of the International Expert meeting on IT-based detection of bioacoustical patterns, Isle of Vilme, Germany.

Gage, S. H., Napoletano, B. M., & Cooper, M. C. (2001). Assessment of ecosystem biodiversity by acoustic diversity indices. *The Journal of the Acoustical Society of America, 109*(5).

Galaxy Zoo. (2010). The Story So Far. Retrieved 7/7/2010, from http://www.galaxyzoo.org/story, http://www.galaxyzoo.org/team

Gasc, A., Sueur, J., Jiguet, F., Devictor, V., Grandcolas, P., Burrow, C., . . . Pavoine, S. (2013). Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? *Ecological Indicators, 25*(0), 279-287. doi: http://dx.doi.org/10.1016/j.ecolind.2012.10.009

Gill, F., & Wright, M. (2006). *Birds of the world: recommended English names*.

Gionis, A., Indyk, P., & Motwani, R. (1999). *Similarity search in high dimensions via hashing.* Paper presented at the VLDB.

Greenwood, J. (2007). Citizens, science and bird conservation. *Journal of Ornithology, 148*(0), 77-124. doi: 10.1007/s10336-007-0239-9

Han, N. C., Muniandy, S. V., & Dayou, J. (2011). Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. *Applied Acoustics*.

Härmä, A. (2003, 6-10 April 2003). *Automatic identification of bird species based on sinusoidal modeling of syllables.* Paper presented at the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03), Hong Kong.

Haykin, S. (1991). *Advances in spectrum analysis and array processing Volume 3*: Prentice-Hall.

Herr, A., Klomp, N. I., & Atkinson, J. S. (1997). Identification of bat echolocation calls using a decision tree classification system. *Complexity International, 4*, 1-9.

Heymann, P., & Garcia-Molina, H. (2006). Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems. Stanford, California: Stanford InfoLab.

Holmes, S. B., McIlwrick, K. A., & Venier, L. A. (2014). Using automated sound recording and analysis to detect bird species-at-risk in southwestern Ontario woodlands. *Wildlife Society Bulletin*, n/a-n/a. doi: 10.1002/wsb.421

Hu, W., Bulusu, N., Chou, C. T., Jha, S., Taylor, A., & Tran, V. N. (2009). Design and evaluation of a hybrid sensor network for cane toad monitoring. *ACM Trans. Sen. Netw., 5*(1), 1-28. doi: 10.1145/1464420.1464424

Huang, K. L., Kanhere, S. S., & Hu, W. (2010). Preserving privacy in participatory sensing systems. *Computer Communications, 33*(11), 1266-1280. doi: 10.1016/j.comcom.2009.08.012

Jankowski, N. W. (2007). Exploring e-Science: An Introduction. *Journal of Computer-Mediated Communication, 12*(2), 549-562.

Keast, A. (1993). Song Structures and Characteristics: Members of a Eucalypt Forest Bird Community Compared. *Emu, 93*(4), 259-268.

Kindt, R., & Coe, R. (2005). *Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies*: World Agroforestry Centre.

Kirschel, A. N., Earl, D. A., Yao, Y., Escobar, I. A., Vilches, E., Vallejo, E. E., & Taylor, C. E. (2009). Using songs to identify individual Mexican antthrush Formicarius moniliger: Comparison of four classification methods. *Bioacoustics, 19*(1-2), 1-20.

Kirschel, A. N. G., Blumstein, D. T., Cohen, R. E., Buermann, W., Smith, T. B., & Slabbekoorn, H. (2009). Birdsong tuned to the environment: green hylia song varies with elevation, tree cover, and noise. *Behavioral Ecology, 20*(5), 1089-1095. doi: 10.1093/beheco/arp101

Kroodsma, D. E., & Miller, E. H. (1996). *Ecology and evolution of acoustic communication in birds*: Comstock Pub.

Kubat, R., DeCamp, P., Roy, B., & Roy, D. (2007). *Totalrecall: visualization and semi-automatic annotation of very large audio-visual corpora*. Paper presented at the 9th international conference on Multimodal interfaces, Nagoya, Aichi, Japan.

Lazarevic, L., Harrison, D., Southee, D., Wade, M., & Osmond, J. (2008). Wind farm and fauna interaction: detecting bird and bat wing beats through cyclic motion analysis. *International Journal of Sustainable Engineering, 1*(1), 60-68.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., . . . Andreescu, D. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society, 389*(3), 1179-1189.

Ma, W. Y., & Manjunath, B. S. (1994, 31 Oct-2 Nov 1994). *Pattern retrieval in image databases based on adaptive signal decomposition.* Paper presented at the 1994 Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers Pacific Grove, California.

Mainwaring, A., Culler, D., Polastre, J., Szewczyk, R., & Anderson, J. (2002). *Wireless sensor networks for habitat monitoring*. Paper presented at the Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications, Atlanta, Georgia, USA. http://portal.acm.org/citation.cfm?id=570751

Marler, P. R., & Slabbekoorn, H. (Eds.). (2004). *Nature's music: the science of birdsong*: Academic Press.

Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). *HT06, tagging paper, taxonomy, Flickr, academic article, to read*. Paper presented at the Proceedings of the seventeenth conference on Hypertext and hypermedia, Odense, Denmark. http://dx.doi.org/10.1145/1149941.1149949

Mason, R., Roe, P., Towsey, M., Jinglan, Z., Gibson, J., & Gage, S. (2008). *Towards an Acoustic Environmental Observatory*. Paper presented at the IEEE Fourth International Conference on eScience, 2008., Indiana

Mathes, A. (2004). Folksonomies-cooperative classification and communication through shared metadata. *Computer Mediated Communication, 47*(10).

McCallum, A. (2010). Birding by ear, visually. *Birding, 42*, 50-63.

McClatchie, S., Thorne, R. E., Grimes, P., & Hanchet, S. (2000). Ground truth and target identification for fisheries acoustics. *Fisheries Research, 47*(2–3), 173-191. doi: http://dx.doi.org/10.1016/S0165-7836(00)00168-5

McIlraith, A. L., & Card, H. C. (1997). Birdsong recognition using backpropagation and multivariate statistics. *IEEE Transactions on Signal Processing, 45*(11), 2740-2748.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1985). *Machine learning: An artificial intelligence approach* (Vol. 1): Morgan Kaufmann.

Mitchell, T. M. (1999). Machine learning and data mining. *Commun. ACM, 42*(11), 30-36. doi: 10.1145/319382.319388

Moore, S. E., Stafford, K. M., Mellinger, D. K., & Hildebrand, J. A. (2006). Listening for Large Whales in the Offshore Waters of Alaska. *Bioscience, 56*(1), 49-55.

National Audubon Society. (2010). Christmas Bird Count. Retrieved 09/07/2010, from http://www.audubon.org/bird/cbc/

Nattkemper, T. W., Twellmann, T., Ritter, H., & Schubert, W. (2003). Human vs. machine: evaluation of fluorescence micrographs. *Computers in biology and medicine, 33*(1), 31-43.

Palialexis, A., Georgakarakos, S., Karakassis, I., Lika, K., & Valavanis, V. (2011). Prediction of marine species distribution from presence–absence acoustic data: comparing the fitting efficiency and the predictive capacity of conventional and novel distribution models. *Hydrobiologia*, 1-26.

Pieretti, N., Farina, A., & Morri, D. (2011). A new methodology to infer the singing activity of an avian community: the Acoustic Complexity Index (ACI). *Ecological Indicators, 11*(3), 868-873.

Planitz, B. M., Roe, P., Sumitomo, J., Towsey, M. W., Williamson, I., & Wimmer, J. (2009). *Listening to nature: techniques for large-scale monitoring of ecosystems using acoustics*. Paper presented at the 3rd eResearch Australasia Conference, 9-13 November 2009,, Novotel Sydney.

Planitz, B. M., Roe, P., Sumitomo, J., Towsey, M. W., Williamson, I., Wimmer, J., & Zhang, J. (2009). *Listening to nature: acoustic monitoring of the environment*. Paper presented at the Microsoft eScience Workshop 2009, 15–17 October 2009, Carnegie Mellon University, Pittsburgh.

Porter, J., Arzberger, P., Braun, H.-W., Bryant, P., Gage, S., Hansen, T., . . . Williams, T. (2005). Wireless Sensor Networks for Ecology. *Bioscience, 55*(7), 561-572. doi: 10.1641/0006-3568(2005)055[0561:wsnfe]2.0.co;2

Potamitis, I., Ntalampiras, S., Jahn, O., & Riede, K. (2014). Automatic bird sound detection in long real-field recordings: Applications and tools. *Applied Acoustics, 80*(0), 1-9. doi: http://dx.doi.org/10.1016/j.apacoust.2014.01.001

Reddy, S., Shilton, K., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2008). Evaluating participation and performance in participatory sensing. *UrbanSense08, November, 4*.

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., St, . . . Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Commun. ACM, 47*(1), 57-59. doi: 10.1145/962081.962106

Ricci, F., Rokach, L., & Shapira, B. (2011). *Introduction to recommender systems handbook*: Springer.

Rickwood, P., & Taylor, A. (2008). Methods for automatically analyzing humpback song units. *The Journal of the Acoustical Society of America, 123*(3), 1763-1772.

Riede, K. (1993). Monitoring Biodiversity: Analysis of Amazonian Rainforest Sounds. *Ambio, 22*(8), 546-548.

Rusu, A., & Govindaraju, V. (2004, 26-29 Oct. 2004). *Handwritten CAPTCHA: using the difference in the abilities of humans and machines in reading handwritten words.* Paper presented at the Ninth International Workshop on Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004, Tokyo, Japan.

Sayigh, L., Quick, N., Hastie, G., & Tyack, P. (2013). Repeated call types in short-finned pilot whales, Globicephala macrorhynchus. *Marine Mammal Science, 29*(2), 312-324. doi: 10.1111/j.1748-7692.2012.00577.x

Scharenborg, O. (2007). Reaching over the gap: A review of efforts to link human and automatic speech recognition research. *Speech Communication, 49*(5), 336-347. doi: http://dx.doi.org/10.1016/j.specom.2007.01.009

Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). *Methods and metrics for cold-start recommendations*. Paper presented at the Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, Tampere, Finland.

Schroeter, R., Hunter, J., Guerin, J., Khan, I., & Henderson, M. (2006, Dec. 2006). *A Synchronous Multimedia Annotation System for Secure Collaboratories.* Paper presented at the Second IEEE International Conference on e-Science and Grid Computing.

Shamir, L., Yerby, C., Simpson, R., von Benda-Beckmann, A. M., Tyack, P., Samarra, F., . . . Wallin, J. (2014). Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. *The Journal of the Acoustical Society of America, 135*(2), 953-962. doi: doi:http://dx.doi.org/10.1121/1.4861348

Shneiderman, B. (2003). *Designing The User Interface: Strategies for Effective Human-Computer Interaction, 4/e (New Edition)*. Reading, Mass: Pearson Education India.

Simpson, K., & Day, N. (1996). *The Princeton field guide to the birds of Australia*: Princeton University Press.

Skowronski, M. D., & Harris, J. G. (2006). Acoustic detection and classification of microchiroptera using machine learning: Lessons learned from automatic speech recognition. *The Journal of the Acoustical Society of America, 119*(3), 1817. doi: http://dx.doi.org/10.1121/1.2166948

Slater, P. J. B. (2003). Fifty years of bird song research: a case study in animal behaviour. *Animal Behaviour, 65*(4), 633-639. doi: 10.1006/anbe.2003.2051

Sokal, R. R. (1974). Classification: purposes, principles, progress, prospects. *Science, 185*(4157), 1115-1123.

Somervuo, P., Harma, A., & Fagerlund, S. (2006). Parametric Representations of Bird Sounds for Automatic Species Recognition. *IEEE Transactions on Audio, Speech, and Language Processing, 14*(6), 2252-2263.

Sroka, J. J., & Braida, L. D. (2005). Human and machine consonant recognition. *Speech Communication, 45*(4), 401-423. doi: http://dx.doi.org/10.1016/j.specom.2004.11.009

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*(3736), 652-654.

Sueur, J., Pavoine, S., Hamerlynck, O., & Duvail, S. (2008). Rapid Acoustic Survey for Biodiversity Appraisal. *PLoS ONE, 3*(12), e4065.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation, 142*(10), 2282-2292.

Sullivan, R. (2009, Jun/Jul 2009). Citizen science BREAKS NEW GROUND. *ECOS Magazine,* 10-13.

Tachibana, R. O., Oosugi, N., & Okanoya, K. (2014). Semi-Automatic Classification of Birdsong Elements Using a Linear Support Vector Machine. *PLoS ONE, 9*(3), e92584. doi: 10.1371/journal.pone.0092584

Taylor, A., Watson, G., Grigg, G., & McCallum, H. (1996, 5-7 August 1996). *Monitoring Frog Communities: An Application of Machine Learning.* Paper presented at the Proceedings of The Eighth Annual Conference on Innovative Applications of Artificial Intelligence, Portland, Oregon.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., . . . Williams, S. E. (2004). Extinction risk from climate change. *Nature, 427*(6970), 145-148. doi: 10.1038/nature02121

Towsey, M., Parsons, S., & Sueur, J. (2014). Ecology and acoustics at a large scale. *Ecological Informatics*(0). doi: http://dx.doi.org/10.1016/j.ecoinf.2014.02.002

Towsey, M., Planitz, B., Nantes, A., Wimmer, J., & Roe, P. (2012). A toolbox for animal call recognition. *Bioacoustics, 21*(2), 107-125. doi: 10.1080/09524622.2011.648753

Towsey, M., Wimmer, J., Williamson, I., & Roe, P. (2014). The use of acoustic indices to determine avian species richness in audio-recordings of the environment. *Ecological Informatics, 21*(0), 110-119. doi: http://dx.doi.org/10.1016/j.ecoinf.2013.11.007

Towsey, M., Zhang, L., Cottman-Fields, M., Wimmer, J., Zhang, J., & Roe, P. (2014). *Visualization of long-duration acoustic recordings of the environment*. Paper presented at the The International Conference on Computational Science, Cairns, Australia.

Towsey, M. W., & Planitz, B. (2010). *Technical Report: Acoustic Analysis of the Natural Environment*.

Towsey, M. W., Wimmer, J., Williamson, I., Roe, P., & Grace, P. (2012). The calculation of acoustic indices to characterise acoustic recordings of the environment. QUT ePrints, Brisbane, Australia.

Tucker, D., Gage, S., Williamson, I., & Fuller, S. (2014). Linking ecological condition and the soundscape in fragmented Australian forests. *Landscape Ecology, 29*(4), 745-758. doi: 10.1007/s10980-014-0015-1

Tyagi, V., & Wellekens, C. (2005, March 18-23, 2005). *On desensitizing the Mel-Cepstrum to spurious spectral components for Robust Speech Recognition.* Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005 (ICASSP '05). .

Vander Wal, T. (2007). Folksonomy. *online posting, Feb.* 2014 from http://vanderwal.net/folksonomy.html

Vaseghi, S. V. (2008). *Advanced digital signal processing and noise reduction*: John Wiley & Sons.

Versi, E. (1992). "Gold standard" is an appropriate term. *BMJ, 305*(6846), 187.

Villanueva-Rivera, L. J., & Pijanowski, B. C. (2012). Pumilio: A Web-Based Management System for Ecological Recordings. *Bulletin of the Ecological Society of America, 93*(1), 71-81. doi: 10.1890/0012-9623-93.1.71

Waddle, J. H., Thigpen, T. F., & Glorioso, B. M. (2009). Efficacy of Automatic Vocalization Recognition Software for Anuran Monitoring. *Herpetological Conservation and Biology, 4*(3), 384-388.

Wang, A. (2006). The Shazam music recognition service. *Commun. ACM, 49*(8), 44-48. doi: 10.1145/1145287.1145312

Wildlife Acoustics. (2011). Song Scope Product Page. Retrieved 23/05/2011, from http://www.wildlifeacoustics.com/songscope.php

Wimmer, J., Towsey, M., Planitz, B., Williamson, I., & Roe, P. (2012). Analysing environmental acoustic data through collaboration and automation. *Future Generation Computer Systems*. doi: 10.1016/j.future.2012.03.004

Wimmer, J., Towsey, M., Planitz, B., Williamson, I., & Roe, P. (2013). Analysing environmental acoustic data through collaboration and automation. *Future Generation Computer Systems, 29*(2), 560-568. doi: http://dx.doi.org/10.1016/j.future.2012.03.004

Wimmer, J., Towsey, M., Roe, P., & Williamson, I. (2013). Sampling environmental acoustic recordings to determine bird species richness. *Ecological applications*. doi: 10.1890/12-2088.1

Wolf, K. (2009). *Bird Song Recognition through Spectrogram Processing and Labeling*. Unpublished manuscript, University of Minnesota. Retrieved from http://www.tc.umn.edu/~wolfx265/DREU/project/final_report/final_report.pdf

Wood, C., Sullivan, B., Iliff, M., Fink, D., & Kelling, S. (2011). eBird: Engaging Birders in Science and Conservation. *PLoS Biol, 9*(12), e1001220. doi: 10.1371/journal.pbio.1001220

Xeno-canto Foundation. (2012). Frequently Asked Questions. Retrieved 06/08/2013, from http://www.xeno-canto.org/FAQ.php

Xeno-canto Foundation. (2013). Sharing bird sounds from around the world. Retrieved 06/08/2013, from http://www.xeno-canto.org

Xu, Z., Fu, Y., Mao, J., & Su, D. (2006). *Towards the semantic web: Collaborative tag suggestions.* Paper presented at the Collaborative web tagging workshop at WWW2006, Edinburgh, Scotland.

Zezula, P., Amato, G., Dohnal, V., & Batko, M. (2006). *Similarity search: the metric space approach* (Vol. 32): Springer.

Zhang, J., Huang, K., Cottman-Fields, M., Truskinger, A., Roe, P., Duan, S., . . . Wimmer, J. (2013, 3-5 Dec. 2013). *Managing and Analysing Big Audio Data for Environmental Monitoring.* Paper presented at the 2013 IEEE 16th International Conference on Computational Science and Engineering (CSE), Sydney, Australia.

Appendices

Appendix A – QUT Thesis by Publication Regulations

> *14.1.1 The Queensland University of Technology permits the presentation of theses for the degree of Doctor of Philosophy in the format of published and/or submitted papers, where such papers have been published, accepted or submitted during the period of candidature; and where the quality of such papers is appropriate to PhD-level research. For the purpose of this Regulation, papers are defined as journal articles, book chapters, conference papers and other forms of written scholarly works which are subject to a process of peer review similar to that of refereed journals. Creative works are not included in this definition. Such works are dealt with in Regulation 15 (below).*
>
> *14.1.2 Papers submitted as a PhD thesis must be closely related in terms of subject matter and form a cohesive research narrative.*
>
> *14.1.3 The minimum number of papers and/or manuscripts is normally three. However, in some disciplines a larger number of papers is required to meet the expectations of scope and quality commensurate with PhD-level research. At least one paper must have been published, accepted, or be undergoing revision following refereeing. The faculty research committee is the source of appropriate advice to PhD candidates with respect to disciplinary norms in these matters.*
>
> *14.2.3 Where the papers have multiple authorship, the candidate must be principal author on at least two of the three papers and have written permission of the co-authors*

Appendix B – Ethics Application Approval

Note: the following approval application approves the thesis author as well. The ethics application has primary and secondary researcher roles. I was listed as a secondary researcher.

From: QUT Research Ethics Unit
Sent: Monday, 18 June 2012 4:48 PM
To: Mark Cottman-Fields
Cc: QUT Research Ethics Unit; Paul Roe
Subject: Ethics Application Approval – 1200000307

Dear Mr Mark Cottman-Fields

Project Title:
Ecosounds – Crowdsourcing the analysis of recorded environmental audio

Approval Number: 1200000307
Clearance Until: 18/06/2015
Ethics Category: Human

This email is to advise that your application has been reviewed by the Chair, University Human Research Ethics Committee, and confirmed as meeting the requirements of the National Statement on Ethical Conduct in Human Research.

Whilst the data collection of your project has received ethical clearance, the decision to commence and authority to commence may be dependent on factors beyond the remit of the ethics review process. For example, your research may need ethics clearance from other organisations or permissions from other organisations to access staff. Therefore the proposed data collection should not commence until you have satisfied these requirements.

If you require a formal approval certificate, please respond via reply email and one will be issued.

Decisions related to low risk ethical review are subject to ratification at the next available Committee meeting. You will only be contacted again in relation to this matter if the Committee raises any additional questions or concerns.

This project has been awarded ethical clearance until 18/06/2015 and a progress report must be submitted for an active ethical clearance at least once every twelve months. Researchers who fail to submit an appropriate progress report may have their ethical clearance revoked and/or the ethical clearances of other projects suspended. When your project has been completed please advise us by email at your earliest convenience.

For information regarding the use of social media in research, please go to:
http://www.research.qut.edu.au/ethics/humans/faqs/index.jsp

For variations, please complete and submit an online variation form:
http://www.research.qut.edu.au/ethics/humans/applications.jsp#amend

Please do not hesitate to contact the unit if you have any queries.

Regards

Janette Lamb on behalf of the Chair UHREC
Research Ethics Unit | Office of Research
Level 4 | 88 Musk Avenue | Kelvin Grove
p: +61 7 3138 5123
e: ethicscontact@qut.edu.au
w: http://www.research.qut.edu.au/ethics/

Appendix C – Participant Information Sheet

Participant Information for QUT Research Project

–- Questionnaire and Website -–

Ecosounds

QUT Ethics Approval Number 1200000307

RESEARCH TEAM

Principal Researcher: Mark-Cottman Fields, Masters Student, QUT

Associate Researcher: Anthony Truskinger, PhD Student, QUT

Supervisor: Professor Paul Roe, QUT

DESCRIPTION

This project is being undertaken as part of a PhD study for Anthony Truskinger and a Masters study for Mark Cottman-Fields.

The purpose of this research is to investigate ways for interested volunteers to interact with audio recordings of environmental sounds, with a focus on sounds made by animals such as bird, frog, and koala calls. The recordings are from areas with little to no human activity. The two main goals of the research are to provide ecologists with information about the animals present in the area the sound recordings were made, and create a website that enables people to listen to and label audio in an easy and straightforward way.

You are invited to participate in this project if you have an interest in the project. The research team is looking for anyone who is interested in listening to audio recordings of the natural environment, and willing to indicate interesting sounds or suggest the name of the animal that made a call. For example, people with bird watching experience, people who camp or hike regularly, people who enjoy listening to the sounds of nature, or those who live outside the city.

You will need access to a computer, an Internet connection, and some way to listen to audio, such as speakers or a pair of headphones.

PARTICIPATION

Your participation in this project is entirely voluntary. If you do agree to participate, you can withdraw from the project without comment or penalty. Any personally identifiable information already obtained from you will be destroyed. However, other contributed data will be made anonymous and kept. Your decision to participate (or not), will in no way impact upon your current or future relationship with QUT.

To participate, complete and submit a questionnaire (on the website), or create an account on the website.

You can withdraw by not submitting a questionnaire. You can stop and discard an incomplete or complete questionnaire before you submit it. Once a questionnaire is submitted, it is not possible to withdraw it, as the questionnaire is anonymous.

You can withdraw by deleting your account on the website. If you delete your account, information you have entered in your personal profile will be deleted. Other contributions you may have made, such as adding tags, will be made anonymous and kept.

Participation will involve interaction with a number of web pages and completing questionnaires about your experience with these web pages. Each web page is designed in a slightly different way to investigate the most effective way to collect data and help the environment. Some of these web pages will be more like a traditional question/answer survey (some with Likert Scales – e.g. strongly agree to strongly disagree and some short answer questions).

These web pages involve completing tasks that will take varying amounts of time. However each task is designed to be small (no more than a minute) and you can choose to stop participating at any time.

The tasks will include activities like listening to audio on a web page, and pressing a button to indicate an interesting sound. Other input may include suggestions for the animal that made the sound. Tasks may show a visualisation of recorded audio. An example activity might be drawing a rectangle around a part of the visualisation to indicate an area of interest (an interesting sound you might be hearing in the recorded audio).

The questionnaires will include questions about the user interfaces, your reaction to them, and experience with them. For example “Did you understand the tasks you were asked to complete?” and “Did the visualisation accurately represent the sound you heard?”

Your participation in this project may include content created by other participants on the website. This might occur through viewing the history of another participant’s activity on the website, or as part of a list of suggested tags. The website may include competitive elements, such as a display of participants ranked by the amount of work each has completed up to that point in time.

If you agree to participate you do not have to answer any question(s) or complete any task(s) that you are uncomfortable with.

EXPECTED BENEFITS

It is expected that this project will not directly benefit you. However, it may benefit you in that you may learn to recognise a range of animal calls as well as exposing you to other people interested in the environment and listening to animal calls. As part of this project you may receive training in identifying particular calls. This would take the form of informative pages on the website that give examples of different species, or tips on how to associate spectrograms (visualisations of audio) with underlying audio.

When you complete a task, you may receive feedback on your performance in that task. This feedback will be calculated based on data collected by other participants on the website. Again, the feedback is provided so that you may know when you have correctly identified an animal call. In this way we hope you will be able to apply this knowledge for other recreational activities.

This project aims to benefit environmental research, particularly research into animal populations. This can include studies of changes in particular environmental regions and evaluations of proposed development works on the environment. In this way any participation on your behalf will go towards understanding the environment.

RISKS

There are no risks beyond normal day-to-day living associated with your participation in this project.

PRIVACY AND CONFIDENTIALITY

All comments, responses, and personally identifiable user contributed data will be treated confidentially.
The actions that you perform to complete tasks on the website may be monitored, such as the buttons you click and the identifier of the recorded audio that is loaded. This monitoring may include any personally identifiable information. The monitoring is only for evaluating and improving the website, and will only be available to the research team.
Any data collected as part of this project will be stored securely as per QUT's management of research data policy.
Please note that non-identifiable data collected in this project may be used as comparative data in future projects.
CONSENT TO PARTICIPATE
There are two main ways for you to confirm your consent to participate in this project.
Signing up and creating an account on the website, or clicking 'I Accept' or 'Submit' button at the bottom of the online Participation Consent form on the website is accepted as an indication of your consent to participate in this project.
Submitting a completed online questionnaire is accepted as an indication of your consent to participate in that questionnaire. It does not indicate your consent to participate in other aspects of the project.
If you are involved in a discovery or some other notable event, we may ask you for permission to publically use your username, real name or both. This is entirely optional.
QUESTIONS / FURTHER INFORMATION ABOUT THE PROJECT
If have any questions or require any further information please contact one of the research team members below.

Mark Cottman-Fields - Masters Student
Science and Engineering Faculty
Queensland University of Technology
3138 9340
m.cottman-fields@student.qut.edu.au

Anthony Truskinger – PhD Student
Science and Engineering Faculty
Queensland University of Technology
3138 9340
anthony.truskinger@student.qut.edu.au

Prof Paul Roe – Supervisor
Science and Engineering Faculty
Queensland University of Technology
p.roe@qut.edu.au

CONCERNS / COMPLAINTS REGARDING THE CONDUCT OF THE PROJECT

QUT is committed to research integrity and the ethical conduct of research projects. However, if you do have any concerns or complaints about the ethical conduct of the project you may contact the QUT Research Ethics Unit on 3138 5123 or email ethicscontact@qut.edu.au. The QUT Research Ethics Unit is not connected with the research project and can facilitate a resolution to your concern in an impartial manner. Thank you for helping with this research project. Please note the url of this information sheet for future reference.

Appendix D – Additional Suggestion Tool Results

File	Dataset	Analysis Name	EndFrequency	StartFrequency	TagDuration	TimeOfDay	Inverted AUC	AUC	1 : Accuracy	5 : Accuracy	10 : Accuracy	25 : Accuracy	50 : Accuracy	1 : Sensitivity	5 : Sensitivity	10 : Sensitivity	25 : Sensitivity	50 : Sensitivity	Time Taken	Peak Memory	Time
2012-04-23 16_14_30 BasicGrouped.xlsx	FullSet	BasicGrouped	A	B	C		0.10	0.90	0.16	0.36	0.47	0.64	0.75	0.16	0.36	0.47	0.64	0.75	00:00:03.2621866	195,944,448	3.262
2012-04-23 16_14_34 BasicGroupedAnti.xlsx	FullSet	BasicGroupedAnti	A	B	C		0.38	0.62	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.04	00:00:02.6851536	200,101,888	2.685
2012-04-23 16_14_37 ZScoreGrouped.xlsx	FullSet	ZScoreGrouped	A	B	C		0.03	0.97	0.19	0.47	0.64	0.84	0.97	0.19	0.47	0.64	0.84	0.97	00:00:08.0394599	690,110,464	8.039
2012-04-23 16_14_45 ZScoreGroupedAnti.xlsx	FullSet	ZScoreGroupedAnti	A	B	C		0.23	0.77	0.00	0.01	0.03	0.06	0.16	0.00	0.01	0.03	0.06	0.16	00:00:08.5234875	690,110,464	8.523
2012-04-23 16_14_54 ZScoreGroupedSingleFix.xlsx	FullSet	ZScoreGroupedSingleFix	A	B	C		0.04	0.96	0.13	0.40	0.58	0.80	0.95	0.13	0.40	0.58	0.80	0.95	00:00:08.9825138	690,110,464	8.983
2012-04-23 16_15_03 ZScoreGroupedAntiSingleFix.xlsx	FullSet	ZScoreGroupedAntiSingleFix	A	B	C		0.24	0.76	0.00	0.00	0.00	0.03	0.16	0.00	0.00	0.00	0.03	0.16	00:00:09.4105382	690,110,464	9.411
2012-04-23 16_15_13 Basic.xlsx	FullSet	Basic	A	B	C		0.00	1.00	0.43	0.64	0.71	0.78	0.84	0.43	0.64	0.71	0.78	0.84	00:14:02.5471910	23,857,868,800	842.547
2012-04-23 16_29_19 BasicAnti.xlsx	FullSet	BasicAnti	A	B	C		0.01	0.99	0.02	0.08	0.15	0.29	0.43	0.02	0.08	0.15	0.29	0.43	00:14:10.6896566	23,857,868,800	850.690
2012-04-23 16_43_34 BasicGrouped.xlsx	FullSet	BasicGrouped		B	C		0.12	0.88	0.07	0.26	0.40	0.59	0.71	0.07	0.26	0.40	0.59	0.71	00:00:02.9981715	195,944,448	2.998
2012-04-23 16_43_37 BasicGroupedAnti.xlsx	FullSet	BasicGroupedAnti		B	C		0.44	0.56	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.02	00:00:02.6441512	200,101,888	2.644
2012-04-23 16_43_40 ZScoreGrouped.xlsx	FullSet	ZScoreGrouped		B	C		0.05	0.95	0.07	0.30	0.46	0.70	0.87	0.07	0.30	0.46	0.70	0.87	00:00:08.0834624	690,110,464	8.083
2012-04-23 16_43_49 ZScoreGroupedAnti.xlsx	FullSet	ZScoreGroupedAnti		B	C		0.20	0.80	0.01	0.01	0.02	0.05	0.22	0.01	0.01	0.02	0.05	0.22	00:00:08.4274821	690,110,464	8.427
2012-04-23 16_43_57 ZScoreGroupedSingleFix.xlsx	FullSet	ZScoreGroupedSingleFix		B	C		0.07	0.93	0.05	0.23	0.39	0.64	0.81	0.05	0.23	0.39	0.64	0.81	00:00:08.3524777	690,110,464	8.352
2012-04-23 16_44_06 ZScoreGroupedAntiSingleFix.xlsx	FullSet	ZScoreGroupedAntiSingleFix		B	C		0.27	0.73	0.00	0.00	0.00	0.10	0.18	0.00	0.00	0.00	0.10	0.18	00:00:08.4454830	690,110,464	8.445
2012-04-23 16_44_15 Basic.xlsx	FullSet	Basic		B	C		0.01	0.99	0.27	0.50	0.58	0.68	0.74	0.27	0.50	0.58	0.68	0.74	00:12:42.6906234	23,857,868,800	762.691
2012-04-23 16_57_02 BasicAnti.xlsx	FullSet	BasicAnti		B	C		0.01	0.99	0.02	0.10	0.17	0.31	0.43	0.02	0.10	0.17	0.31	0.43	00:12:46.4438380	23,857,868,800	766.444

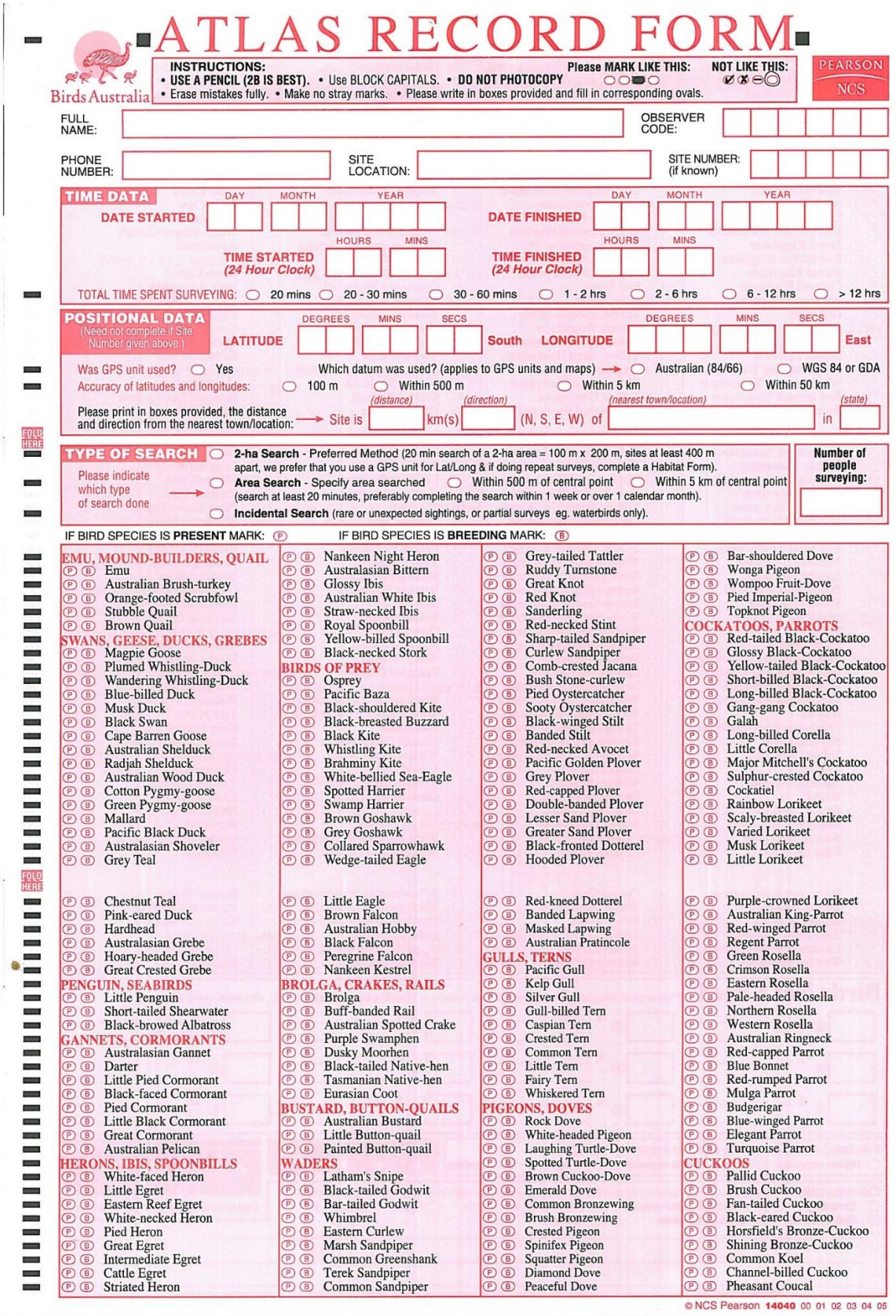
2012-04-23 17_09_51 BasicGrouped.xlsx	FullSet	BasicGrouped	A	C	0.12	0.88	0.06	0.22	0.32	0.51	0.67	0.06	0.22	0.32	0.51	0.67	00:00:03.0641752	195,944,448	3.064
2012-04-23 17_09_55 BasicGroupedAnti.xlsx	FullSet	BasicGroupedAnti	A	C	0.41	0.59	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	00:00:02.9041661	200,101,888	2.904
2012-04-23 17_09_58 ZScoreGrouped.xlsx	FullSet	ZScoreGrouped	A	C	0.06	0.94	0.04	0.21	0.36	0.62	0.89	0.04	0.21	0.36	0.62	0.89	00:00:07.9534549	690,110,464	7.953
2012-04-23 17_10_06 ZScoreGroupedAnti.xlsx	FullSet	ZScoreGroupedAnti	A	C	0.22	0.78	0.00	0.00	0.01	0.11	0.27	0.00	0.00	0.01	0.11	0.27	00:00:08.8945088	690,110,464	8.895
2012-04-23 17_10_16 ZScoreGroupedSingleFix.xlsx	FullSet	ZScoreGroupedSingleFix	A	C	0.08	0.92	0.04	0.17	0.30	0.54	0.78	0.04	0.17	0.30	0.54	0.78	00:00:08.3774792	690,110,464	8.377
2012-04-23 17_10_24 ZScoreGroupedAntiSingleFix.xlsx	FullSet	ZScoreGroupedAntiSingleFix	A	C	0.29	0.71	0.00	0.00	0.00	0.03	0.16	0.00	0.00	0.00	0.03	0.16	00:00:08.5384884	690,110,464	8.538
2012-04-23 17_10_33 Basic.xlsx	FullSet	Basic	A	C	0.00	1.00	0.22	0.47	0.58	0.69	0.76	0.22	0.47	0.58	0.69	0.76	00:12:43.0636447	23,857,868,800	763.064
2012-04-23 17_23_20 BasicAnti.xlsx	FullSet	BasicAnti	A	C	0.02	0.98	0.02	0.09	0.16	0.31	0.44	0.02	0.09	0.16	0.31	0.44	00:13:01.0276723	23,857,868,800	781.028
2012-04-23 17_36_24 BasicGrouped.xlsx	FullSet	BasicGrouped	A	B	0.12	0.88	0.10	0.27	0.39	0.58	0.73	0.10	0.27	0.39	0.58	0.73	00:00:02.7161554	195,944,448	2.716
2012-04-23 17_36_27 BasicGroupedAnti.xlsx	FullSet	BasicGroupedAnti	A	B	0.41	0.59	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.02	00:00:02.6181497	200,101,888	2.618
2012-04-23 17_36_30 ZScoreGrouped.xlsx	FullSet	ZScoreGrouped	A	B	0.05	0.95	0.12	0.33	0.48	0.72	0.94	0.12	0.33	0.48	0.72	0.94	00:00:08.8755076	690,110,464	8.876
2012-04-23 17_36_39 ZScoreGroupedAnti.xlsx	FullSet	ZScoreGroupedAnti	A	B	0.26	0.74	0.00	0.00	0.01	0.07	0.13	0.00	0.00	0.01	0.07	0.13	00:00:08.5934915	690,110,464	8.593
2012-04-23 17_36_48 ZScoreGroupedSingleFix.xlsx	FullSet	ZScoreGroupedSingleFix	A	B	0.06	0.94	0.07	0.25	0.39	0.64	0.85	0.07	0.25	0.39	0.64	0.85	00:00:08.4724846	690,110,464	8.472
2012-04-23 17_36_57 ZScoreGroupedAntiSingleFix.xlsx	FullSet	ZScoreGroupedAntiSingleFix	A	B	0.33	0.67	0.00	0.02	0.02	0.04	0.07	0.00	0.02	0.02	0.04	0.07	00:00:08.8795078	690,110,464	8.880
2012-04-23 17_37_06 Basic.xlsx	FullSet	Basic	A	B	0.00	1.00	0.23	0.50	0.59	0.69	0.75	0.23	0.50	0.59	0.69	0.75	00:12:42.8426321	23,857,868,800	762.843
2012-04-23 17_49_52 BasicAnti.xlsx	FullSet	BasicAnti	A	B	0.01	0.99	0.02	0.08	0.15	0.29	0.43	0.02	0.08	0.15	0.29	0.43	00:13:03.7408274	23,857,868,800	783.741
2012-04-23 18_02_59 BasicGrouped.xlsx	FullSet	BasicGrouped		C	0.21	0.79	0.01	0.05	0.10	0.27	0.47	0.01	0.05	0.10	0.27	0.47	00:00:02.8101607	195,944,448	2.810
2012-04-23 18_03_03 BasicGroupedAnti.xlsx	FullSet	BasicGroupedAnti		C	0.49	0.51	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.01	0.02	00:00:02.6761531	200,101,888	2.676
2012-04-23 18_03_06 ZScoreGrouped.xlsx	FullSet	ZScoreGrouped		C	0.14	0.86	0.02	0.07	0.14	0.32	0.54	0.02	0.07	0.14	0.32	0.54	00:00:07.7164413	690,110,464	7.716
2012-04-23 18_03_14 ZScoreGroupedAnti.xlsx	FullSet	ZScoreGroupedAnti		C	0.23	0.77	0.00	0.00	0.01	0.12	0.20	0.00	0.00	0.01	0.12	0.20	00:00:07.2734160	690,110,464	7.273
2012-04-23 18_03_21 ZScoreGroupedSingleFix.xlsx	FullSet	ZScoreGroupedSingleFix		C	0.19	0.81	0.01	0.05	0.10	0.23	0.43	0.01	0.05	0.10	0.23	0.43	00:00:08.4384827	690,110,464	8.438
2012-04-23 18_03_30 ZScoreGroupedAntiSingleFix.xlsx	FullSet	ZScoreGroupedAntiSingleFix		C	0.30	0.70	0.00	0.00	0.01	0.03	0.11	0.00	0.00	0.01	0.03	0.11	00:00:08.0474603	690,110,464	8.047

2012-04-23 18_03_39 Basic.xlsx	FullSet	Basic			C		0.01	0.99	0.07	0.24	0.35	0.49	0.60	0.07	0.24	0.35	0.49	0.60	00:12:06.2975418	23,857,868,800	726.298
2012-04-23 18_15_49 BasicAnti.xlsx	FullSet	BasicAnti			C		0.01	0.99	0.02	0.08	0.14	0.29	0.43	0.02	0.08	0.14	0.29	0.43	00:12:14.1049884	23,857,868,800	734.105
2012-04-23 18_28_06 BasicGrouped.xlsx	FullSet	BasicGrouped		B			0.16	0.84	0.01	0.09	0.19	0.40	0.59	0.01	0.09	0.19	0.40	0.59	00:00:03.2191841	195,944,448	3.219
2012-04-23 18_28_10 BasicGroupedAnti.xlsx	FullSet	BasicGroupedAnti		B			0.44	0.56	0.00	0.00	0.00	0.02	0.03	0.00	0.00	0.00	0.02	0.03	00:00:03.2201842	200,101,888	3.220
2012-04-23 18_28_13 ZScoreGrouped.xlsx	FullSet	ZScoreGrouped		B			0.10	0.90	0.02	0.13	0.24	0.49	0.68	0.02	0.13	0.24	0.49	0.68	00:00:09.2925315	690,110,464	9.293
2012-04-23 18_28_23 ZScoreGroupedAnti.xlsx	FullSet	ZScoreGroupedAnti		B			0.27	0.73	0.00	0.01	0.02	0.04	0.14	0.00	0.01	0.02	0.04	0.14	00:00:11.5526608	690,110,464	11.553
2012-04-23 18_28_35 ZScoreGroupedSingleFix.xlsx	FullSet	ZScoreGroupedSingleFix		B			0.13	0.87	0.01	0.07	0.15	0.39	0.59	0.01	0.07	0.15	0.39	0.59	00:00:10.4485977	690,110,464	10.449
2012-04-23 18_28_46 ZScoreGroupedAntiSingleFix.xlsx	FullSet	ZScoreGroupedAntiSingleFix		B			0.36	0.64	0.00	0.00	0.00	0.02	0.06	0.00	0.00	0.00	0.02	0.06	00:00:10.5996062	690,110,464	10.600
2012-04-23 18_28_57 Basic.xlsx	FullSet	Basic		B			0.01	0.99	0.10	0.31	0.42	0.57	0.66	0.10	0.31	0.42	0.57	0.66	00:13:24.5280163	23,857,868,800	804.528
2012-04-23 18_42_26 BasicAnti.xlsx	FullSet	BasicAnti		B			0.01	0.99	0.01	0.06	0.16	0.30	0.44	0.01	0.06	0.16	0.30	0.44	00:11:48.5005239	23,857,868,800	708.501
2012-04-23 18_54_18 BasicGrouped.xlsx	FullSet	BasicGrouped	A				0.18	0.82	0.02	0.09	0.16	0.33	0.53	0.02	0.09	0.16	0.33	0.53	00:00:03.2911883	195,944,448	3.291
2012-04-23 18_54_22 BasicGroupedAnti.xlsx	FullSet	BasicGroupedAnti	A				0.42	0.58	0.00	0.00	0.01	0.02	0.06	0.00	0.00	0.01	0.02	0.06	00:00:03.2871881	200,101,888	3.287
2012-04-23 18_54_26 ZScoreGrouped.xlsx	FullSet	ZScoreGrouped	A				0.10	0.90	0.02	0.09	0.18	0.38	0.67	0.02	0.09	0.18	0.38	0.67	00:00:10.8706217	690,110,464	10.871
2012-04-23 18_54_38 ZScoreGroupedAnti.xlsx	FullSet	ZScoreGroupedAnti	A				0.26	0.74	0.00	0.01	0.01	0.05	0.14	0.00	0.01	0.01	0.05	0.14	00:00:09.7135555	690,110,464	9.714
2012-04-23 18_54_48 ZScoreGroupedSingleFix.xlsx	FullSet	ZScoreGroupedSingleFix	A				0.15	0.85	0.02	0.07	0.13	0.28	0.49	0.02	0.07	0.13	0.28	0.49	00:00:09.5165443	690,110,464	9.517
2012-04-23 18_55_03 ZScoreGroupedAntiSingleFix.xlsx	FullSet	ZScoreGroupedAntiSingleFix	A				0.36	0.64	0.00	0.00	0.00	0.01	0.03	0.00	0.00	0.00	0.01	0.03	00:00:10.7016121	690,110,464	10.702
2012-04-23 18_55_14 Basic.xlsx	FullSet	Basic	A				0.01	0.99	0.08	0.29	0.40	0.54	0.64	0.08	0.29	0.40	0.54	0.64	00:12:47.8929209	23,857,868,800	767.893
2012-04-23 19_08_05 BasicAnti.xlsx	FullSet	BasicAnti	A				0.02	0.98	0.02	0.07	0.14	0.30	0.42	0.02	0.07	0.14	0.30	0.42	00:14:14.0578493	23,857,868,800	854.058
2012-04-24 14_47_09 BasicGrouped.xlsx	FullSet	BasicGrouped	A	B	C	D	0.10	0.90	0.08	0.29	0.42	0.62	0.75	0.08	0.29	0.42	0.62	0.75	00:00:22.3130000	287,920,128	22.313
2012-04-24 14_47_32 BasicGrouped.xlsx	FullSet	BasicGrouped		B	C	D	0.13	0.87	0.02	0.09	0.21	0.48	0.67	0.02	0.09	0.21	0.48	0.67	00:00:19.3920000	292,507,648	19.392
2012-04-24 14_47_52 BasicGrouped.xlsx	FullSet	BasicGrouped	A		C	D	0.13	0.87	0.01	0.09	0.20	0.47	0.67	0.01	0.09	0.20	0.47	0.67	00:00:19.3220000	292,507,648	19.322
2012-04-24 14_48_12 BasicGrouped.xlsx	FullSet	BasicGrouped	A	B	C		0.10	0.90	0.16	0.36	0.47	0.64	0.75	0.16	0.36	0.47	0.64	0.75	00:00:19.3590000	292,507,648	19.359

2012-04-24 14_48_31 BasicGrouped.xlsx	FullSet	BasicGrouped	A	B		D	0.13	0.87	0.01	0.12	0.25	0.49	0.71	0.01	0.12	0.25	0.49	0.71	00:00:20.0400000	292,507,648	20.040
2012-04-24 14_48_52 BasicGrouped.xlsx	FullSet	BasicGrouped			C	D	0.27	0.73	0.00	0.01	0.02	0.08	0.30	0.00	0.01	0.02	0.08	0.30	00:00:19.7660000	292,507,648	19.766
2012-04-24 14_49_12 BasicGrouped.xlsx	FullSet	BasicGrouped		B	C		0.12	0.88	0.07	0.26	0.40	0.59	0.71	0.07	0.26	0.40	0.59	0.71	00:00:19.7880000	292,507,648	19.788
2012-04-24 14_49_32 BasicGrouped.xlsx	FullSet	BasicGrouped		B		D	0.22	0.78	0.00	0.01	0.02	0.12	0.38	0.00	0.01	0.02	0.12	0.38	00:00:19.8110000	292,507,648	19.811
2012-04-24 14_49_53 BasicGrouped.xlsx	FullSet	BasicGrouped	A		C		0.12	0.88	0.06	0.22	0.32	0.51	0.67	0.06	0.22	0.32	0.51	0.67	00:00:20.2150000	292,507,648	20.215
2012-04-24 14_50_13 BasicGrouped.xlsx	FullSet	BasicGrouped	A			D	0.20	0.80	0.00	0.02	0.05	0.13	0.40	0.00	0.02	0.05	0.13	0.40	00:00:20.2490000	292,507,648	20.249
2012-04-24 14_50_34 BasicGrouped.xlsx	FullSet	BasicGrouped	A	B			0.12	0.88	0.10	0.27	0.39	0.58	0.73	0.10	0.27	0.39	0.58	0.73	00:00:19.5390000	292,507,648	19.539
2012-04-24 14_50_54 BasicGrouped.xlsx	FullSet	BasicGrouped			C		0.21	0.79	0.01	0.05	0.10	0.27	0.47	0.01	0.05	0.10	0.27	0.47	00:00:17.8770000	292,507,648	17.877
2012-04-24 14_51_12 BasicGrouped.xlsx	FullSet	BasicGrouped				D	0.79	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	00:00:18.2370000	292,507,648	18.237
2012-04-24 14_51_30 BasicGrouped.xlsx	FullSet	BasicGrouped		B			0.16	0.84	0.01	0.09	0.19	0.40	0.59	0.01	0.09	0.19	0.40	0.59	00:00:18.4760000	292,507,648	18.476
2012-04-24 14_51_49 BasicGrouped.xlsx	FullSet	BasicGrouped	A				0.18	0.82	0.02	0.09	0.16	0.33	0.53	0.02	0.09	0.16	0.33	0.53	00:00:19.3340000	292,507,648	19.334
2012-06-06 12_09_37 GlobalZScore.xlsx	FullSet	GlobalZScore	A	B	C		0.09	0.91	0.17	0.38	0.49	0.66	0.76	0.17	0.38	0.49	0.66	0.76	00:00:24.9124910	322,785,280	24.912
2012-06-06 12_10_03 GlobalZScore.xlsx	FullSet	GlobalZScore		B	C		0.11	0.89	0.07	0.26	0.40	0.59	0.72	0.07	0.26	0.40	0.59	0.72	00:00:20.9110909	326,111,232	20.911
2012-06-06 12_10_25 GlobalZScore.xlsx	FullSet	GlobalZScore	A		C		0.12	0.88	0.06	0.22	0.32	0.52	0.68	0.06	0.22	0.32	0.52	0.68	00:00:17.9767975	326,111,232	17.977
2012-06-06 12_10_43 GlobalZScore.xlsx	FullSet	GlobalZScore	A	B			0.11	0.89	0.10	0.28	0.41	0.61	0.74	0.10	0.28	0.41	0.61	0.74	00:00:18.4138412	326,111,232	18.414
2012-06-06 12_11_02 GlobalZScore.xlsx	FullSet	GlobalZScore			C		0.21	0.79	0.01	0.05	0.10	0.27	0.47	0.01	0.05	0.10	0.27	0.47	00:00:17.6947693	326,111,232	17.695
2012-06-06 12_11_20 GlobalZScore.xlsx	FullSet	GlobalZScore		B			0.16	0.84	0.01	0.09	0.19	0.40	0.59	0.01	0.09	0.19	0.40	0.59	00:00:17.2087207	326,111,232	17.209
2012-06-06 12_11_37 GlobalZScore.xlsx	FullSet	GlobalZScore	A				0.18	0.82	0.02	0.09	0.16	0.33	0.53	0.02	0.09	0.16	0.33	0.53	00:00:18.6708669	326,111,232	18.671
2012-07-02 16_53_50 GlobalZScoreAnti.xlsx	FullSet	GlobalZScoreAnti	A	B	C		0.37	0.63	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	00:00:05.3993089	295,378,944	5.399
2012-07-02 15_27_05 Basic-ReferenceOnly.xlsx	ReferenceTags	Basic	A	B	C		0.21	0.79	0.08	0.24	0.33	0.45	0.56	0.08	0.25	0.33	0.45	0.56	00:17:17.2813291	225,550,336	1037.281
2012-07-02 16_05_48 GlobalZScore.xlsx	ReferenceTags	GlobalZScore	A	B	C		0.22	0.78	0.07	0.20	0.27	0.38	0.48	0.09	0.24	0.33	0.46	0.58	00:00:03.8622209	188,059,648	3.862
2012-06-03 20_14_57 BasicGrouped.xlsx	SmallScaleSet	BasicGrouped	A	B	C		0.14	0.86	0.22	0.49	0.59	0.69	0.76	0.22	0.49	0.59	0.69	0.76	00:00:18.9690850	96,067,584	18.969

2012-06-03 20_15_16 BasicGrouped.xlsx	SmallScaleSet	BasicGrouped	B C	0.15	0.85	0.16	0.41	0.55	0.70	0.77	0.16	0.41	0.55	0.70	0.77	00:00:01.0590606	99,721,216	1.059
2012-06-03 20_15_17 BasicGrouped.xlsx	SmallScaleSet	BasicGrouped	A C	0.16	0.84	0.11	0.33	0.45	0.62	0.73	0.11	0.33	0.45	0.62	0.73	00:00:01.3000744	99,721,216	1.300
2012-06-03 20_15_19 BasicGrouped.xlsx	SmallScaleSet	BasicGrouped	A B	0.17	0.83	0.13	0.37	0.52	0.69	0.77	0.13	0.37	0.52	0.69	0.77	00:00:01.3190754	99,721,216	1.319
2012-06-03 20_15_20 BasicGrouped.xlsx	SmallScaleSet	BasicGrouped	C	0.22	0.78	0.02	0.10	0.22	0.44	0.67	0.02	0.10	0.22	0.44	0.67	00:00:01.3050747	99,721,216	1.305
2012-06-03 20_15_22 BasicGrouped.xlsx	SmallScaleSet	BasicGrouped	B	0.20	0.80	0.06	0.22	0.35	0.58	0.75	0.06	0.22	0.35	0.58	0.75	00:00:01.3030745	99,721,216	1.303
2012-06-03 20_15_23 BasicGrouped.xlsx	SmallScaleSet	BasicGrouped	A	0.22	0.78	0.04	0.16	0.27	0.50	0.70	0.04	0.16	0.27	0.50	0.70	00:00:01.3120751	99,721,216	1.312
2012-06-26 12_06_35 GlobalZScore.xlsx	SmallScaleSet	GlobalZScore	A B C	0.13	0.87	0.24	0.48	0.58	0.67	0.74	0.24	0.50	0.60	0.70	0.77	00:00:07.2042795	99,078,144	7.204
2012-06-26 12_16_29 GlobalZScoreAnti.xlsx	SmallScaleSet	GlobalZScoreAnti	A B C	0.45	0.55	0.00	0.00	0.01	0.04	0.15	0.00	0.00	0.01	0.04	0.15	00:00:05.0070000	84,062,208	5.007

Appendix E – ATLAS Record Form



ATLAS RECORD FORM

Birds Australia

INSTRUCTIONS:
- **USE A PENCIL (2B IS BEST).** • Use BLOCK CAPITALS. • **DO NOT PHOTOCOPY**
- Erase mistakes fully. • Make no stray marks. • Please write in boxes provided and fill in corresponding ovals.

Please MARK LIKE THIS: ● NOT LIKE THIS: ✓ ✗ ⊖ ○

PEARSON NCS

FULL NAME: | OBSERVER CODE: | PHONE NUMBER: | SITE LOCATION: | SITE NUMBER: (if known)

TIME DATA

DATE STARTED (DAY / MONTH / YEAR) | DATE FINISHED (DAY / MONTH / YEAR)

TIME STARTED (24 Hour Clock) (HOURS / MINS) | TIME FINISHED (24 Hour Clock) (HOURS / MINS)

TOTAL TIME SPENT SURVEYING: ○ 20 mins ○ 20 - 30 mins ○ 30 - 60 mins ○ 1 - 2 hrs ○ 2 - 6 hrs ○ 6 - 12 hrs ○ > 12 hrs

POSITIONAL DATA (Need not complete if Site Number given above.)

LATITUDE (DEGREES / MINS / SECS) South | LONGITUDE (DEGREES / MINS / SECS) East

Was GPS unit used? ○ Yes | Which datum was used? (applies to GPS units and maps) → ○ Australian (84/66) ○ WGS 84 or GDA

Accuracy of latitudes and longitudes: ○ 100 m ○ Within 500 m ○ Within 5 km ○ Within 50 km

Please print in boxes provided, the distance and direction from the nearest town/location: → Site is (distance) km(s) (direction) (N, S, E, W) of (nearest town/location) in (state)

TYPE OF SEARCH

Please indicate which type of search done →

○ **2-ha Search** - Preferred Method (20 min search of a 2-ha area = 100 m x 200 m, sites at least 400 m apart, we prefer that you use a GPS unit for Lat/Long & if doing repeat surveys, complete a Habitat Form).

○ **Area Search** - Specify area searched ○ Within 500 m of central point ○ Within 5 km of central point (search at least 20 minutes, preferably completing the search within 1 week or over 1 calendar month).

○ **Incidental Search** (rare or unexpected sightings, or partial surveys eg. waterbirds only).

Number of people surveying:

IF BIRD SPECIES IS **PRESENT** MARK: (P) IF BIRD SPECIES IS **BREEDING** MARK: (B)

EMU, MOUND-BUILDERS, QUAIL
(P) (B) Emu
(P) (B) Australian Brush-turkey
(P) (B) Orange-footed Scrubfowl
(P) (B) Stubble Quail
(P) (B) Brown Quail

SWANS, GEESE, DUCKS, GREBES
(P) (B) Magpie Goose
(P) (B) Plumed Whistling-Duck
(P) (B) Wandering Whistling-Duck
(P) (B) Blue-billed Duck
(P) (B) Musk Duck
(P) (B) Black Swan
(P) (B) Cape Barren Goose
(P) (B) Australian Shelduck
(P) (B) Radjah Shelduck
(P) (B) Australian Wood Duck
(P) (B) Cotton Pygmy-goose
(P) (B) Green Pygmy-goose
(P) (B) Mallard
(P) (B) Pacific Black Duck
(P) (B) Australasian Shoveler
(P) (B) Grey Teal
(P) (B) Chestnut Teal
(P) (B) Pink-eared Duck
(P) (B) Hardhead
(P) (B) Australasian Grebe
(P) (B) Hoary-headed Grebe
(P) (B) Great Crested Grebe

PENGUIN, SEABIRDS
(P) (B) Little Penguin
(P) (B) Short-tailed Shearwater
(P) (B) Black-browed Albatross

GANNETS, CORMORANTS
(P) (B) Australasian Gannet
(P) (B) Darter
(P) (B) Little Pied Cormorant
(P) (B) Black-faced Cormorant
(P) (B) Pied Cormorant
(P) (B) Little Black Cormorant
(P) (B) Great Cormorant
(P) (B) Australian Pelican

HERONS, IBIS, SPOONBILLS
(P) (B) White-faced Heron
(P) (B) Little Egret
(P) (B) Eastern Reef Egret
(P) (B) White-necked Heron
(P) (B) Pied Heron
(P) (B) Great Egret
(P) (B) Intermediate Egret
(P) (B) Cattle Egret
(P) (B) Striated Heron
(P) (B) Nankeen Night Heron
(P) (B) Australasian Bittern
(P) (B) Glossy Ibis
(P) (B) Australian White Ibis
(P) (B) Straw-necked Ibis
(P) (B) Royal Spoonbill
(P) (B) Yellow-billed Spoonbill
(P) (B) Black-necked Stork

BIRDS OF PREY
(P) (B) Osprey
(P) (B) Pacific Baza
(P) (B) Black-shouldered Kite
(P) (B) Black-breasted Buzzard
(P) (B) Black Kite
(P) (B) Whistling Kite
(P) (B) Brahminy Kite
(P) (B) White-bellied Sea-Eagle
(P) (B) Spotted Harrier
(P) (B) Swamp Harrier
(P) (B) Brown Goshawk
(P) (B) Grey Goshawk
(P) (B) Collared Sparrowhawk
(P) (B) Wedge-tailed Eagle
(P) (B) Little Eagle
(P) (B) Brown Falcon
(P) (B) Australian Hobby
(P) (B) Black Falcon
(P) (B) Peregrine Falcon
(P) (B) Nankeen Kestrel

BROLGA, CRAKES, RAILS
(P) (B) Brolga
(P) (B) Buff-banded Rail
(P) (B) Australian Spotted Crake
(P) (B) Purple Swamphen
(P) (B) Dusky Moorhen
(P) (B) Black-tailed Native-hen
(P) (B) Tasmanian Native-hen
(P) (B) Eurasian Coot

BUSTARD, BUTTON-QUAILS
(P) (B) Australian Bustard
(P) (B) Little Button-quail
(P) (B) Painted Button-quail

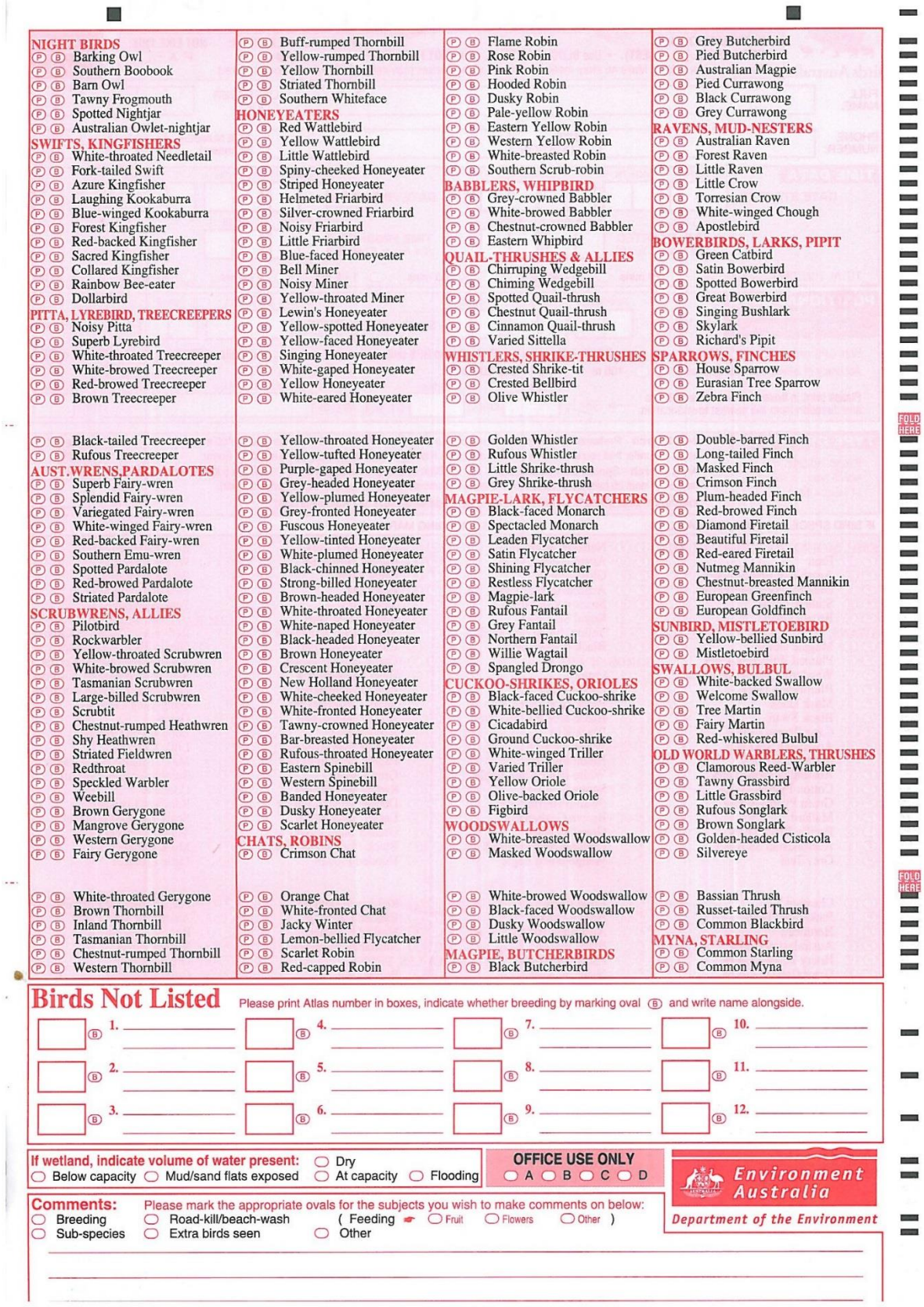
WADERS
(P) (B) Latham's Snipe
(P) (B) Black-tailed Godwit
(P) (B) Bar-tailed Godwit
(P) (B) Whimbrel
(P) (B) Eastern Curlew
(P) (B) Marsh Sandpiper
(P) (B) Common Greenshank
(P) (B) Terek Sandpiper
(P) (B) Common Sandpiper
(P) (B) Grey-tailed Tattler
(P) (B) Ruddy Turnstone
(P) (B) Great Knot
(P) (B) Red Knot
(P) (B) Sanderling
(P) (B) Red-necked Stint
(P) (B) Sharp-tailed Sandpiper
(P) (B) Curlew Sandpiper
(P) (B) Comb-crested Jacana
(P) (B) Bush Stone-curlew
(P) (B) Pied Oystercatcher
(P) (B) Sooty Oystercatcher
(P) (B) Black-winged Stilt
(P) (B) Banded Stilt
(P) (B) Red-necked Avocet
(P) (B) Pacific Golden Plover
(P) (B) Grey Plover
(P) (B) Red-capped Plover
(P) (B) Double-banded Plover
(P) (B) Lesser Sand Plover
(P) (B) Greater Sand Plover
(P) (B) Black-fronted Dotterel
(P) (B) Hooded Plover
(P) (B) Red-kneed Dotterel
(P) (B) Banded Lapwing
(P) (B) Masked Lapwing
(P) (B) Australian Pratincole

GULLS, TERNS
(P) (B) Pacific Gull
(P) (B) Kelp Gull
(P) (B) Silver Gull
(P) (B) Gull-billed Tern
(P) (B) Caspian Tern
(P) (B) Crested Tern
(P) (B) Common Tern
(P) (B) Little Tern
(P) (B) Fairy Tern
(P) (B) Whiskered Tern

PIGEONS, DOVES
(P) (B) Rock Dove
(P) (B) White-headed Pigeon
(P) (B) Laughing Turtle-Dove
(P) (B) Spotted Turtle-Dove
(P) (B) Brown Cuckoo-Dove
(P) (B) Emerald Dove
(P) (B) Common Bronzewing
(P) (B) Brush Bronzewing
(P) (B) Crested Pigeon
(P) (B) Spinifex Pigeon
(P) (B) Squatter Pigeon
(P) (B) Diamond Dove
(P) (B) Peaceful Dove
(P) (B) Bar-shouldered Dove
(P) (B) Wonga Pigeon
(P) (B) Wompoo Fruit-Dove
(P) (B) Pied Imperial-Pigeon
(P) (B) Topknot Pigeon

COCKATOOS, PARROTS
(P) (B) Red-tailed Black-Cockatoo
(P) (B) Glossy Black-Cockatoo
(P) (B) Yellow-tailed Black-Cockatoo
(P) (B) Short-billed Black-Cockatoo
(P) (B) Long-billed Black-Cockatoo
(P) (B) Gang-gang Cockatoo
(P) (B) Galah
(P) (B) Long-billed Corella
(P) (B) Little Corella
(P) (B) Major Mitchell's Cockatoo
(P) (B) Sulphur-crested Cockatoo
(P) (B) Cockatiel
(P) (B) Rainbow Lorikeet
(P) (B) Scaly-breasted Lorikeet
(P) (B) Varied Lorikeet
(P) (B) Musk Lorikeet
(P) (B) Little Lorikeet
(P) (B) Purple-crowned Lorikeet
(P) (B) Australian King-Parrot
(P) (B) Red-winged Parrot
(P) (B) Regent Parrot
(P) (B) Green Rosella
(P) (B) Crimson Rosella
(P) (B) Eastern Rosella
(P) (B) Pale-headed Rosella
(P) (B) Northern Rosella
(P) (B) Western Rosella
(P) (B) Australian Ringneck
(P) (B) Red-capped Parrot
(P) (B) Blue Bonnet
(P) (B) Red-rumped Parrot
(P) (B) Mulga Parrot
(P) (B) Budgerigar
(P) (B) Blue-winged Parrot
(P) (B) Elegant Parrot
(P) (B) Turquoise Parrot

CUCKOOS
(P) (B) Pallid Cuckoo
(P) (B) Brush Cuckoo
(P) (B) Fan-tailed Cuckoo
(P) (B) Black-eared Cuckoo
(P) (B) Horsfield's Bronze-Cuckoo
(P) (B) Shining Bronze-Cuckoo
(P) (B) Common Koel
(P) (B) Channel-billed Cuckoo
(P) (B) Pheasant Coucal



NIGHT BIRDS
(P) (B) Barking Owl
(P) (B) Southern Boobook
(P) (B) Barn Owl
(P) (B) Tawny Frogmouth
(P) (B) Spotted Nightjar
(P) (B) Australian Owlet-nightjar
SWIFTS, KINGFISHERS
(P) (B) White-throated Needletail
(P) (B) Fork-tailed Swift
(P) (B) Azure Kingfisher
(P) (B) Laughing Kookaburra
(P) (B) Blue-winged Kookaburra
(P) (B) Forest Kingfisher
(P) (B) Red-backed Kingfisher
(P) (B) Sacred Kingfisher
(P) (B) Collared Kingfisher
(P) (B) Rainbow Bee-eater
(P) (B) Dollarbird
PITTA, LYREBIRD, TREECREEPERS
(P) (B) Noisy Pitta
(P) (B) Superb Lyrebird
(P) (B) White-throated Treecreeper
(P) (B) White-browed Treecreeper
(P) (B) Red-browed Treecreeper
(P) (B) Brown Treecreeper
(P) (B) Black-tailed Treecreeper
(P) (B) Rufous Treecreeper
AUST.WRENS,PARDALOTES
(P) (B) Superb Fairy-wren
(P) (B) Splendid Fairy-wren
(P) (B) Variegated Fairy-wren
(P) (B) White-winged Fairy-wren
(P) (B) Red-backed Fairy-wren
(P) (B) Southern Emu-wren
(P) (B) Spotted Pardalote
(P) (B) Red-browed Pardalote
(P) (B) Striated Pardalote
SCRUBWRENS, ALLIES
(P) (B) Pilotbird
(P) (B) Rockwarbler
(P) (B) Yellow-throated Scrubwren
(P) (B) White-browed Scrubwren
(P) (B) Tasmanian Scrubwren
(P) (B) Large-billed Scrubwren
(P) (B) Scrubtit
(P) (B) Chestnut-rumped Heathwren
(P) (B) Shy Heathwren
(P) (B) Striated Fieldwren
(P) (B) Redthroat
(P) (B) Speckled Warbler
(P) (B) Weebill
(P) (B) Brown Gerygone
(P) (B) Mangrove Gerygone
(P) (B) Western Gerygone
(P) (B) Fairy Gerygone
(P) (B) White-throated Gerygone
(P) (B) Brown Thornbill
(P) (B) Inland Thornbill
(P) (B) Tasmanian Thornbill
(P) (B) Chestnut-rumped Thornbill
(P) (B) Western Thornbill
(P) (B) Buff-rumped Thornbill
(P) (B) Yellow-rumped Thornbill
(P) (B) Yellow Thornbill
(P) (B) Striated Thornbill
(P) (B) Southern Whiteface
HONEYEATERS
(P) (B) Red Wattlebird
(P) (B) Yellow Wattlebird
(P) (B) Little Wattlebird
(P) (B) Spiny-cheeked Honeyeater
(P) (B) Striped Honeyeater
(P) (B) Helmeted Friarbird
(P) (B) Silver-crowned Friarbird
(P) (B) Noisy Friarbird
(P) (B) Little Friarbird
(P) (B) Blue-faced Honeyeater
(P) (B) Bell Miner
(P) (B) Noisy Miner
(P) (B) Yellow-throated Miner
(P) (B) Lewin's Honeyeater
(P) (B) Yellow-spotted Honeyeater
(P) (B) Yellow-faced Honeyeater
(P) (B) Singing Honeyeater
(P) (B) White-gaped Honeyeater
(P) (B) Yellow Honeyeater
(P) (B) White-eared Honeyeater
(P) (B) Yellow-throated Honeyeater
(P) (B) Yellow-tufted Honeyeater
(P) (B) Purple-gaped Honeyeater
(P) (B) Grey-headed Honeyeater
(P) (B) Yellow-plumed Honeyeater
(P) (B) Grey-fronted Honeyeater
(P) (B) Fuscous Honeyeater
(P) (B) Yellow-tinted Honeyeater
(P) (B) White-plumed Honeyeater
(P) (B) Black-chinned Honeyeater
(P) (B) Strong-billed Honeyeater
(P) (B) Brown-headed Honeyeater
(P) (B) White-throated Honeyeater
(P) (B) White-naped Honeyeater
(P) (B) Black-headed Honeyeater
(P) (B) Brown Honeyeater
(P) (B) Crescent Honeyeater
(P) (B) New Holland Honeyeater
(P) (B) White-cheeked Honeyeater
(P) (B) White-fronted Honeyeater
(P) (B) Tawny-crowned Honeyeater
(P) (B) Bar-breasted Honeyeater
(P) (B) Rufous-throated Honeyeater
(P) (B) Eastern Spinebill
(P) (B) Western Spinebill
(P) (B) Banded Honeyeater
(P) (B) Dusky Honeyeater
(P) (B) Scarlet Honeyeater
CHATS, ROBINS
(P) (B) Crimson Chat
(P) (B) Orange Chat
(P) (B) White-fronted Chat
(P) (B) Jacky Winter
(P) (B) Lemon-bellied Flycatcher
(P) (B) Scarlet Robin
(P) (B) Red-capped Robin
(P) (B) Flame Robin
(P) (B) Rose Robin
(P) (B) Pink Robin
(P) (B) Hooded Robin
(P) (B) Dusky Robin
(P) (B) Pale-yellow Robin
(P) (B) Eastern Yellow Robin
(P) (B) Western Yellow Robin
(P) (B) White-breasted Robin
(P) (B) Southern Scrub-robin
BABBLERS, WHIPBIRD
(P) (B) Grey-crowned Babbler
(P) (B) White-browed Babbler
(P) (B) Chestnut-crowned Babbler
(P) (B) Eastern Whipbird
QUAIL-THRUSHES & ALLIES
(P) (B) Chirruping Wedgebill
(P) (B) Chiming Wedgebill
(P) (B) Spotted Quail-thrush
(P) (B) Chestnut Quail-thrush
(P) (B) Cinnamon Quail-thrush
(P) (B) Varied Sittella
WHISTLERS, SHRIKE-THRUSHES
(P) (B) Crested Shrike-tit
(P) (B) Crested Bellbird
(P) (B) Olive Whistler
(P) (B) Golden Whistler
(P) (B) Rufous Whistler
(P) (B) Little Shrike-thrush
(P) (B) Grey Shrike-thrush
MAGPIE-LARK, FLYCATCHERS
(P) (B) Black-faced Monarch
(P) (B) Spectacled Monarch
(P) (B) Leaden Flycatcher
(P) (B) Satin Flycatcher
(P) (B) Shining Flycatcher
(P) (B) Restless Flycatcher
(P) (B) Magpie-lark
(P) (B) Rufous Fantail
(P) (B) Grey Fantail
(P) (B) Northern Fantail
(P) (B) Willie Wagtail
(P) (B) Spangled Drongo
CUCKOO-SHRIKES, ORIOLES
(P) (B) Black-faced Cuckoo-shrike
(P) (B) White-bellied Cuckoo-shrike
(P) (B) Cicadabird
(P) (B) Ground Cuckoo-shrike
(P) (B) White-winged Triller
(P) (B) Varied Triller
(P) (B) Yellow Oriole
(P) (B) Olive-backed Oriole
(P) (B) Figbird
WOODSWALLOWS
(P) (B) White-breasted Woodswallow
(P) (B) Masked Woodswallow
(P) (B) White-browed Woodswallow
(P) (B) Black-faced Woodswallow
(P) (B) Dusky Woodswallow
(P) (B) Little Woodswallow
MAGPIE, BUTCHERBIRDS
(P) (B) Black Butcherbird
(P) (B) Grey Butcherbird
(P) (B) Pied Butcherbird
(P) (B) Australian Magpie
(P) (B) Pied Currawong
(P) (B) Black Currawong
(P) (B) Grey Currawong
RAVENS, MUD-NESTERS
(P) (B) Australian Raven
(P) (B) Forest Raven
(P) (B) Little Raven
(P) (B) Little Crow
(P) (B) Torresian Crow
(P) (B) White-winged Chough
(P) (B) Apostlebird
BOWERBIRDS, LARKS, PIPIT
(P) (B) Green Catbird
(P) (B) Satin Bowerbird
(P) (B) Spotted Bowerbird
(P) (B) Great Bowerbird
(P) (B) Singing Bushlark
(P) (B) Skylark
(P) (B) Richard's Pipit
SPARROWS, FINCHES
(P) (B) House Sparrow
(P) (B) Eurasian Tree Sparrow
(P) (B) Zebra Finch
(P) (B) Double-barred Finch
(P) (B) Long-tailed Finch
(P) (B) Masked Finch
(P) (B) Crimson Finch
(P) (B) Plum-headed Finch
(P) (B) Red-browed Finch
(P) (B) Diamond Firetail
(P) (B) Beautiful Firetail
(P) (B) Red-eared Firetail
(P) (B) Nutmeg Mannikin
(P) (B) Chestnut-breasted Mannikin
(P) (B) European Greenfinch
(P) (B) European Goldfinch
SUNBIRD, MISTLETOEBIRD
(P) (B) Yellow-bellied Sunbird
(P) (B) Mistletoebird
SWALLOWS, BULBUL
(P) (B) White-backed Swallow
(P) (B) Welcome Swallow
(P) (B) Tree Martin
(P) (B) Fairy Martin
(P) (B) Red-whiskered Bulbul
OLD WORLD WARBLERS, THRUSHES
(P) (B) Clamorous Reed-Warbler
(P) (B) Tawny Grassbird
(P) (B) Little Grassbird
(P) (B) Rufous Songlark
(P) (B) Brown Songlark
(P) (B) Golden-headed Cisticola
(P) (B) Silvereye
(P) (B) Bassian Thrush
(P) (B) Russet-tailed Thrush
(P) (B) Common Blackbird
MYNA, STARLING
(P) (B) Common Starling
(P) (B) Common Myna

FOLD HERE

FOLD HERE

Birds Not Listed
Please print Atlas number in boxes, indicate whether breeding by marking oval (B) and write name alongside.

1. ____ 2. ____ 3. ____
4. ____ 5. ____ 6. ____
7. ____ 8. ____ 9. ____
10. ____ 11. ____ 12. ____

If wetland, indicate volume of water present:
○ Dry ○ Below capacity ○ Mud/sand flats exposed ○ At capacity ○ Flooding

OFFICE USE ONLY
○ A ○ B ○ C ○ D

Environment Australia
Department of the Environment

Comments:
Please mark the appropriate ovals for the subjects you wish to make comments on below:
○ Breeding ○ Sub-species ○ Road-kill/beach-wash ○ Extra birds seen (Feeding ☛ ○ Fruit ○ Flowers ○ Other) ○ Other

Appendix F – Annotation Software Platform Screenshots

The following screenshots are all taken from the current QUT Bioacoustic Software platform. All of these screenshots were obtained from the publically accessible website (http://sensor.mquter.qut.edu.au/) on the 1st of September 2013.

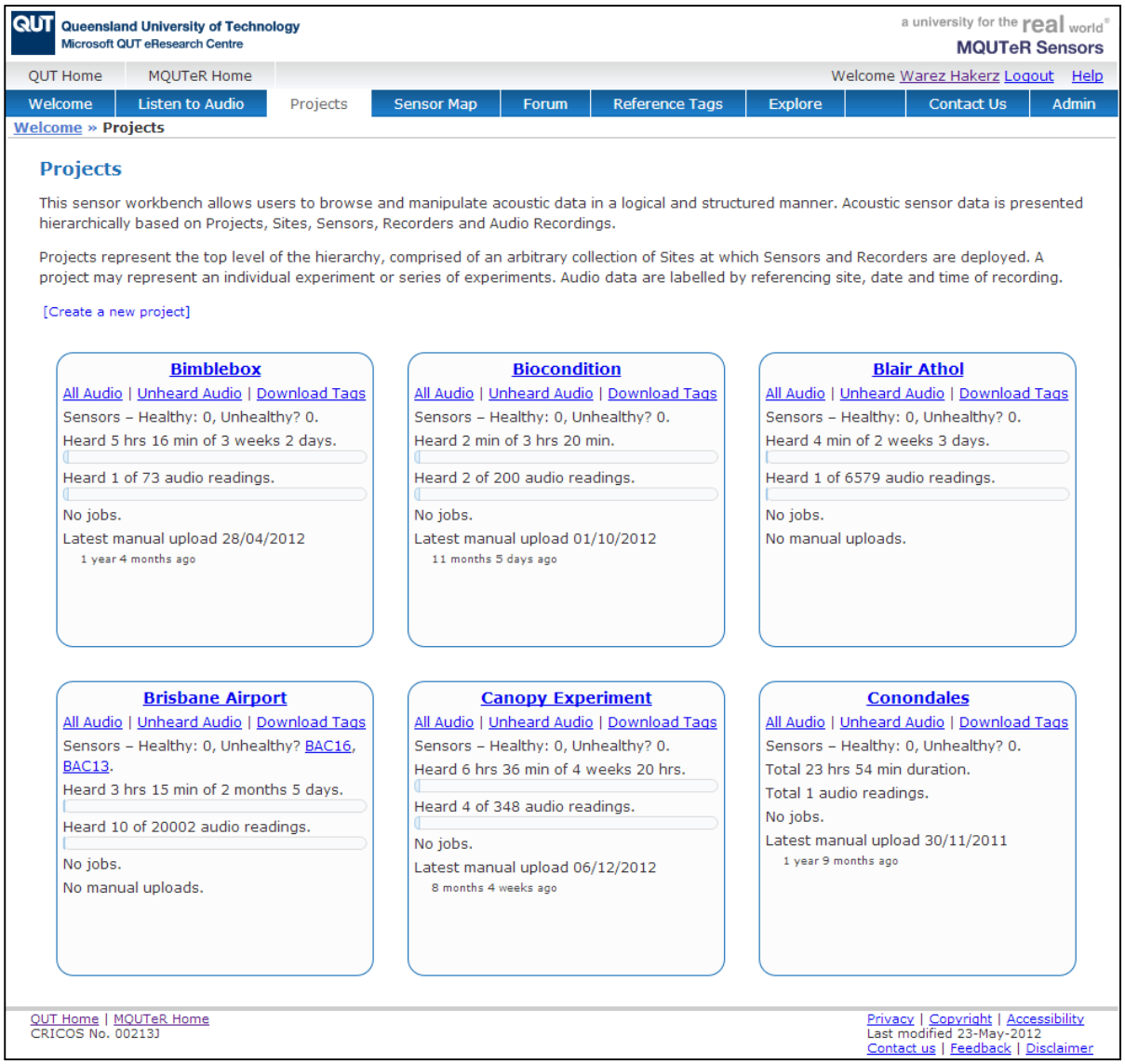


Figure 15 – A screenshot of the Project listing screen

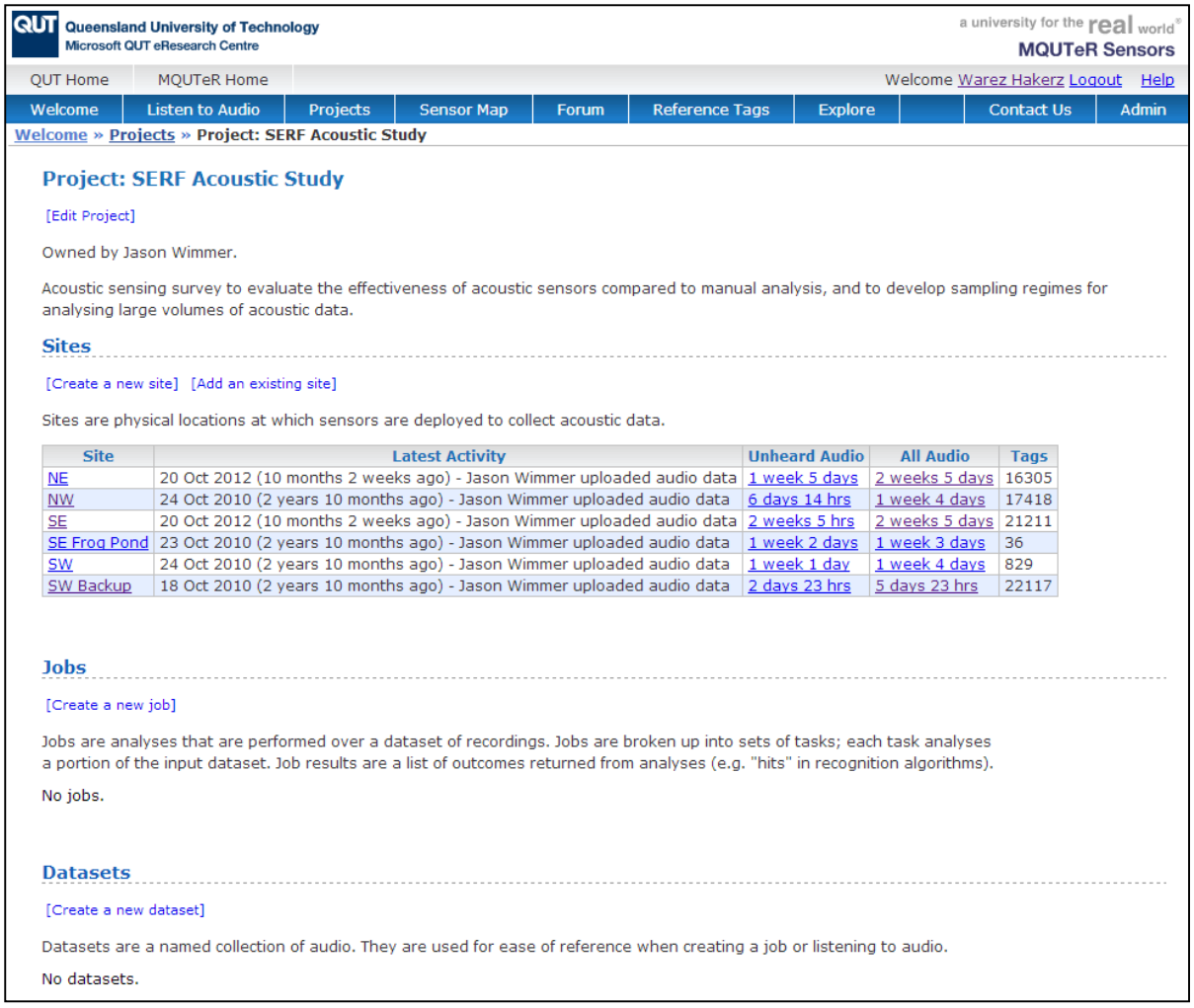


Figure 16 – A screenshot of the Project details page

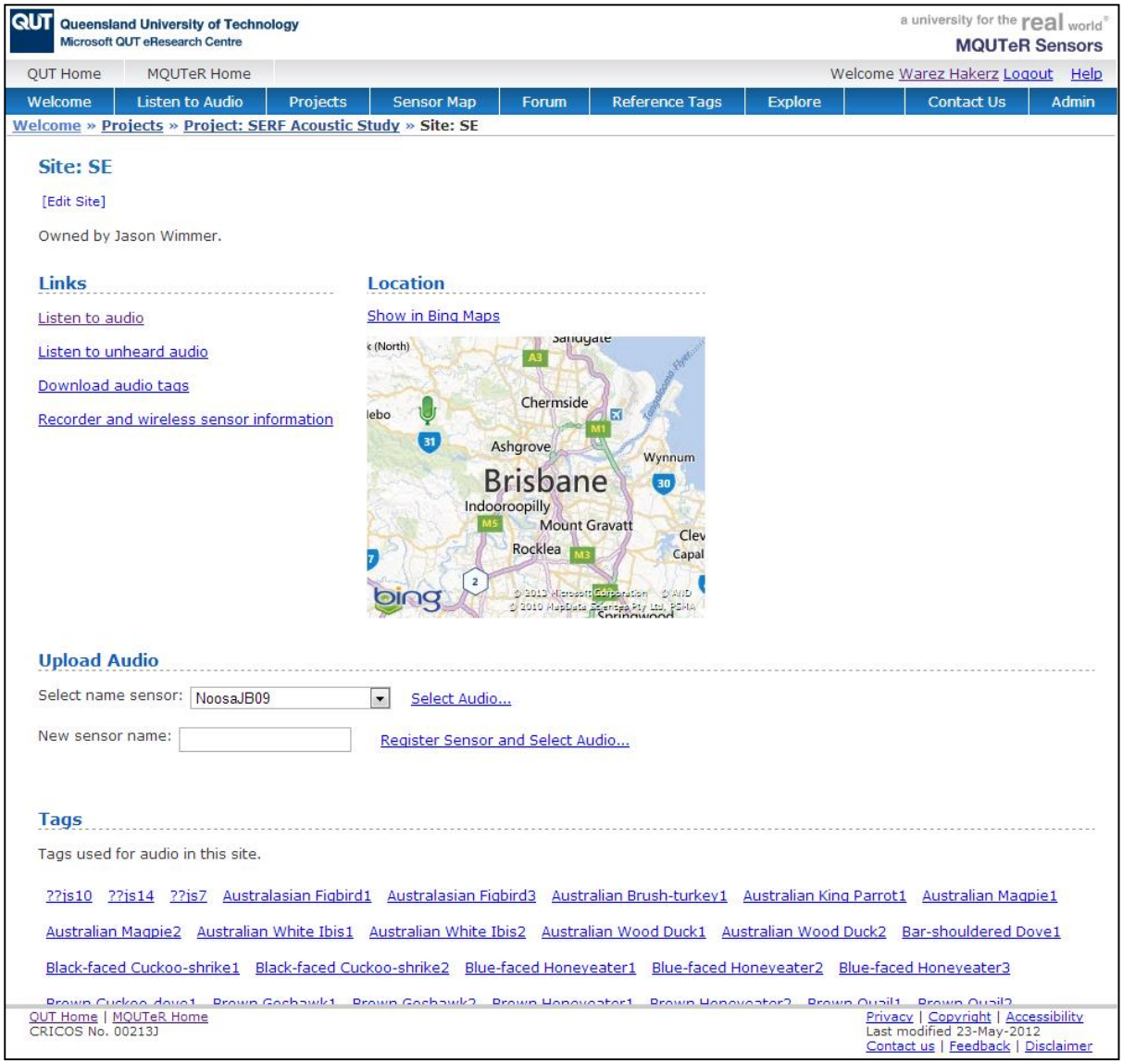


Figure 17 – A screenshot of the Site details page

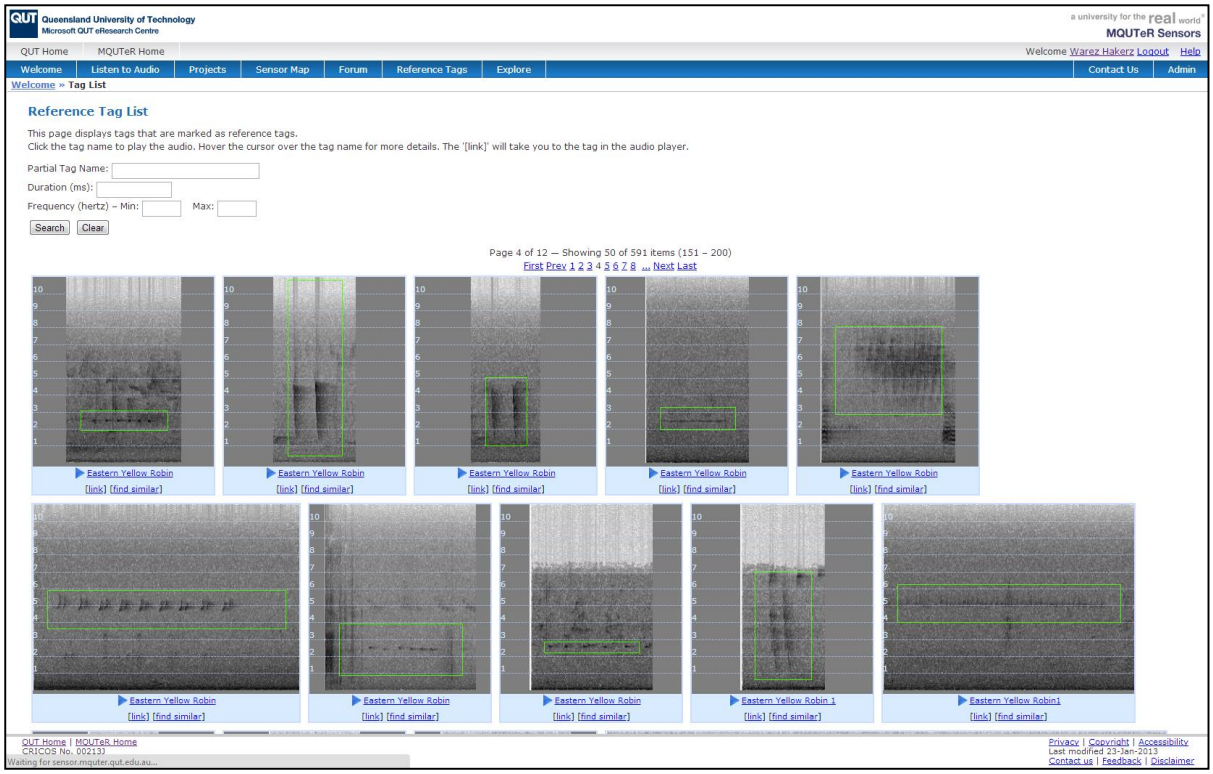


Figure 18 – A screenshot of the Reference Library, used to assist annotators

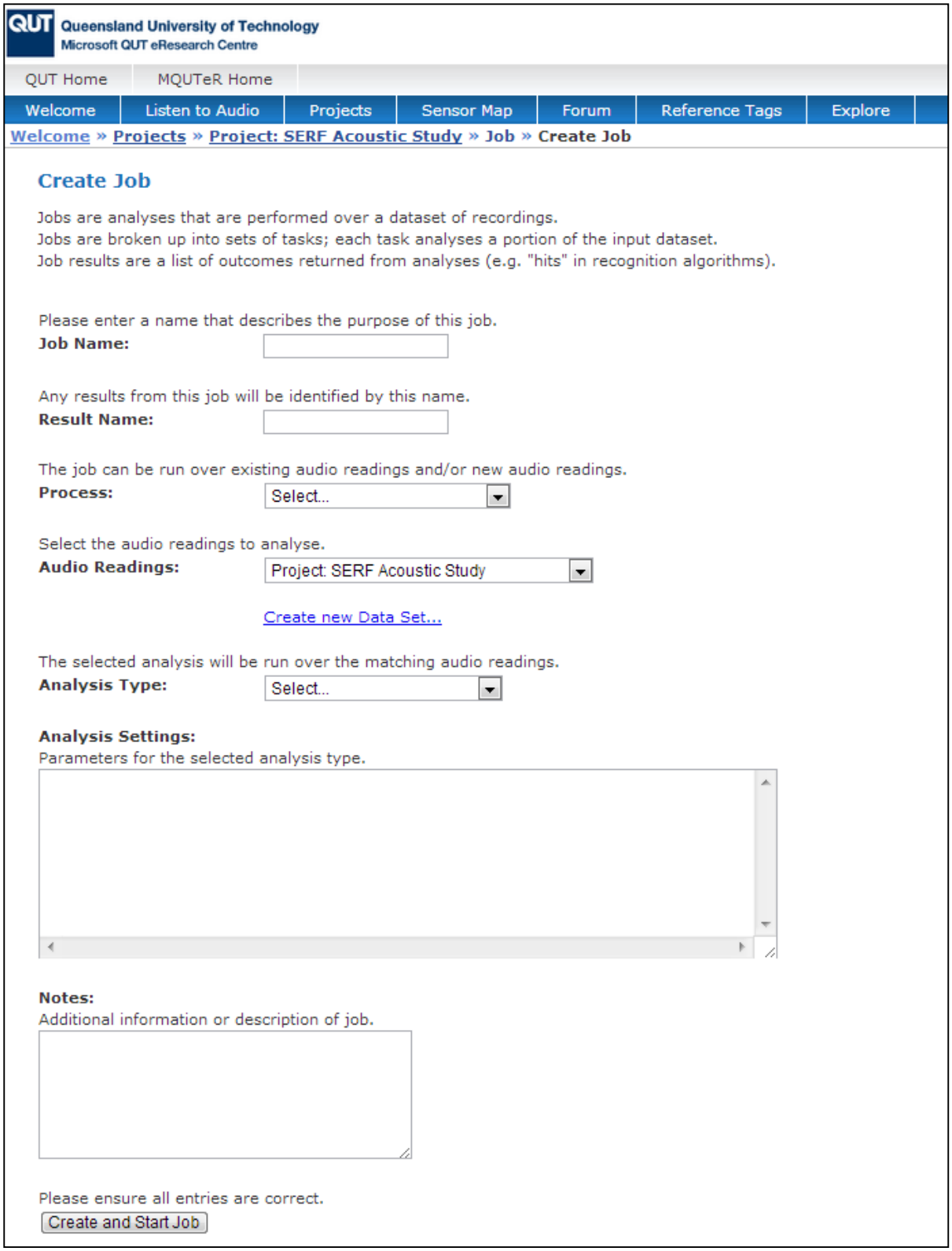


Figure 19 – A screenshot of the Job creation page. This page is used for scheduling the automated analysis of a selected subset of data.

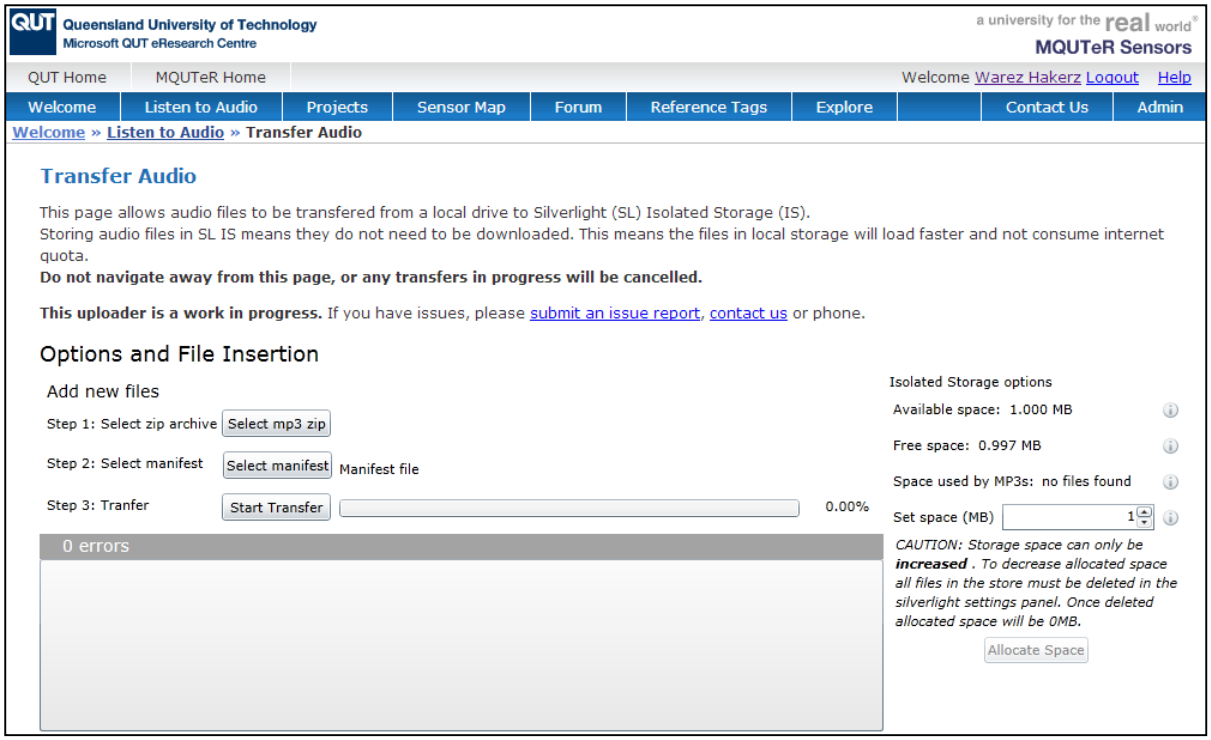


Figure 20 – A screenshot of the audio transfer application used to install media locally to a users' computer (to save on bandwidth).

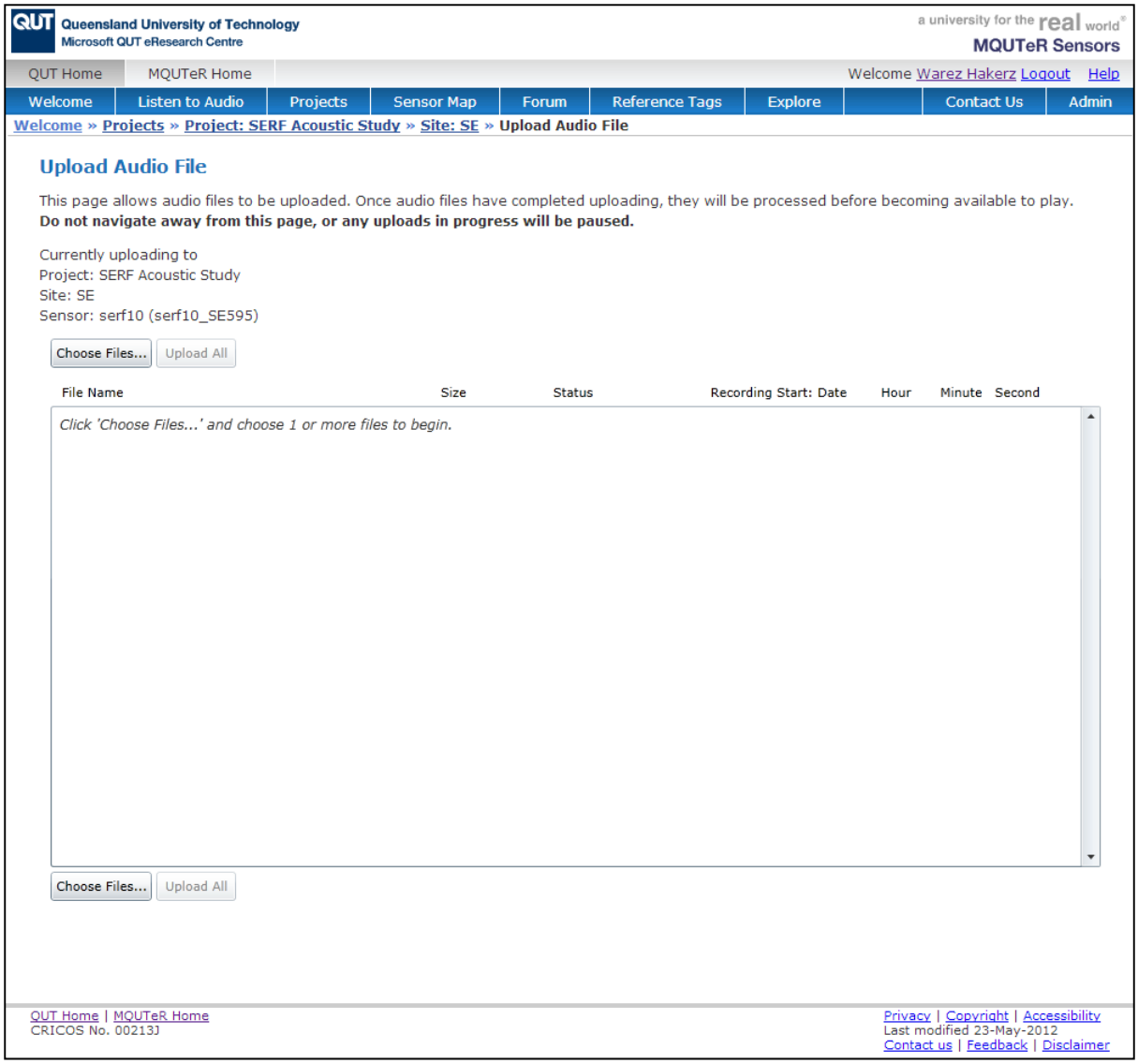


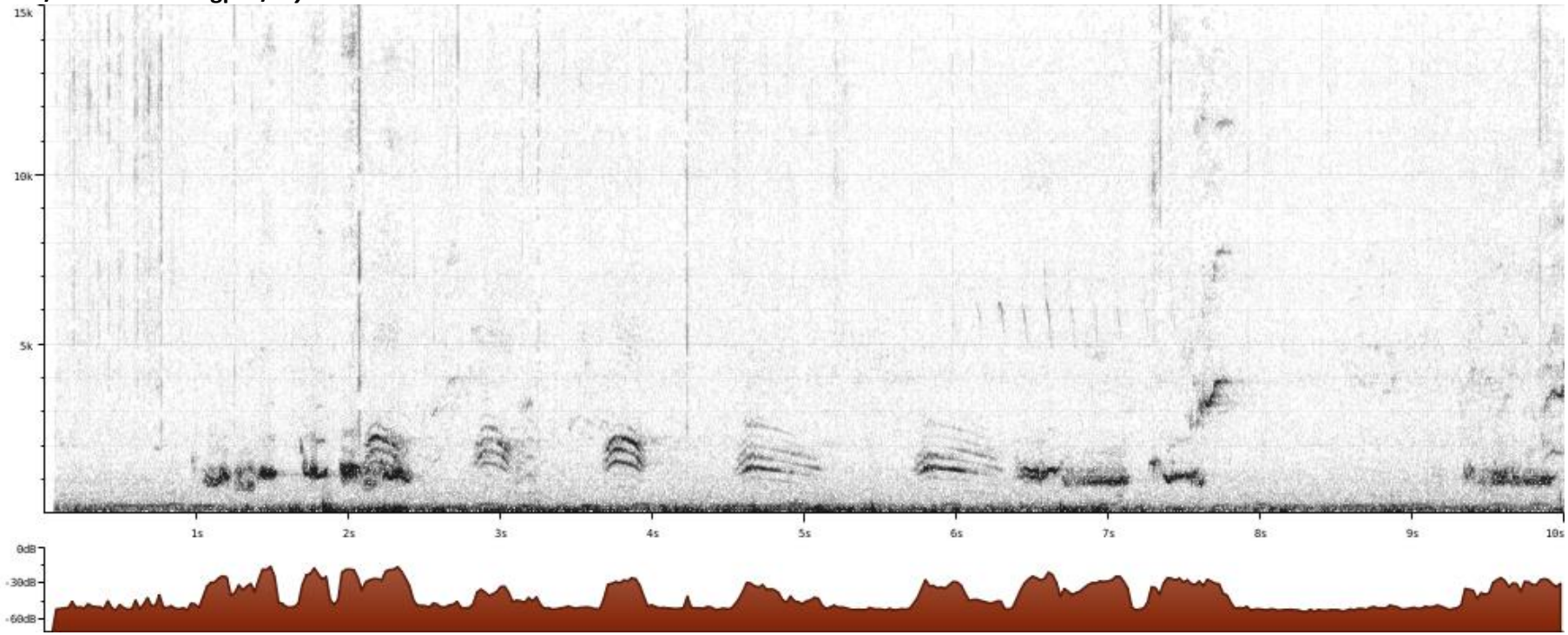
Figure 21 – A screenshot of the bulk audio upload interface

Appendix G – Examples of Faunal Vocalisations

This appendix contains examples of faunal vocalisations. The subset shown are all Australian avian species (with the exception of the human sample at the end) that were recorded near Brisbane, Australia. All recordings were taken from the Xeno-Canto website and reproduced here under the Creative Commons Attribution-NonCommercial-NoDerivatives 2.5 Generic licence.

<<Continued on next page>>

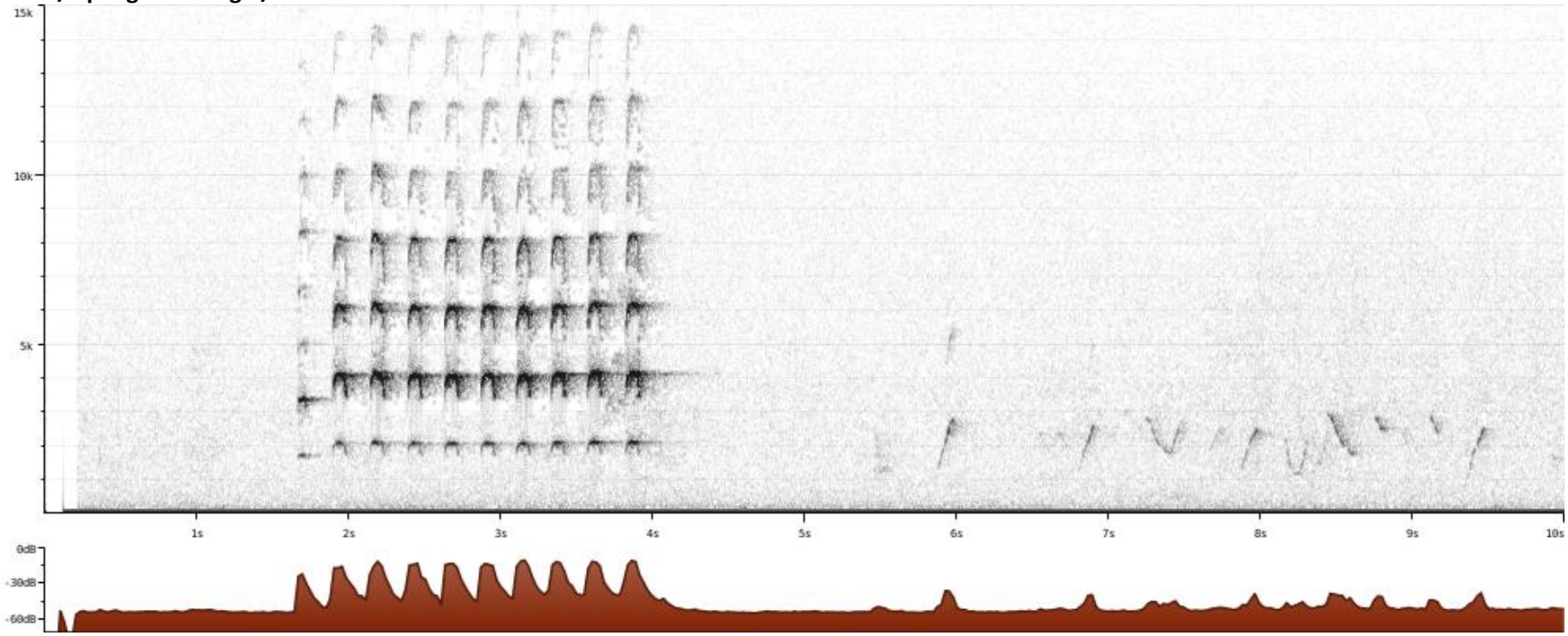
XC63369 / Australian Magpie / *Gymnorhina tibicen*



http://www.xeno-canto.org/63369 / http://www.xeno-canto.org/sounds/uploaded/SILWLBBIFA/ffts/XC63369-large.png

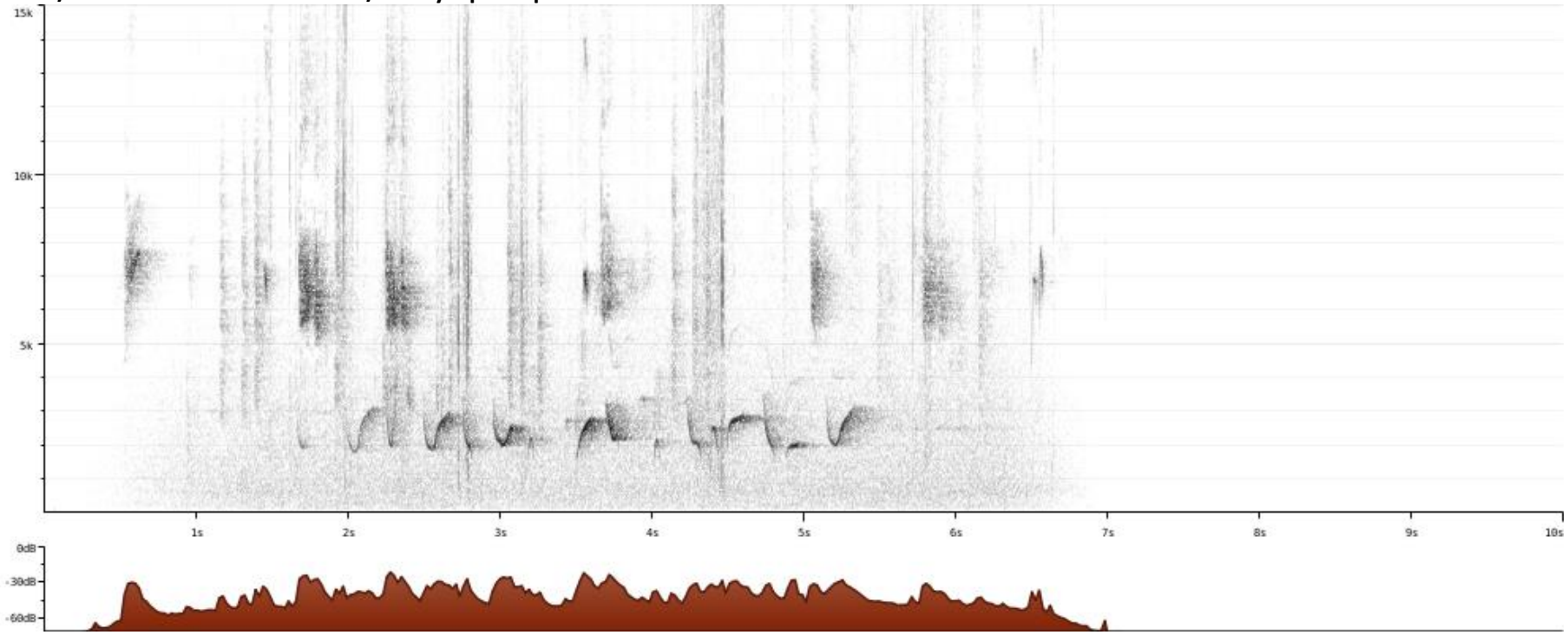
Recordist	**Peter Woodall**
Date	2010-01-17
Time	13:54
Latitude	-27.096
Longitude	153.167
Location	**Buckley's Hole, Bribie Island**
Country	Australia
Elevation	30 m
Background	**Torresian Crow** (*Corvus orru*)

XC101561 / Spangled Drongo / *Dicrurus bracteatus*



http://www.xeno-canto.org/101561 / http://www.xeno-canto.org/sounds/uploaded/NRUIEQXSTF/ffts/XC101561-large.png Recordist	**Judith Lattaway**
Date	2012-05-21
Time	1700
Latitude	-27.1334
Longitude	153.0167
Location	**Burpengary Qld**
Country	Australia
Elevation	11 m
Background	none

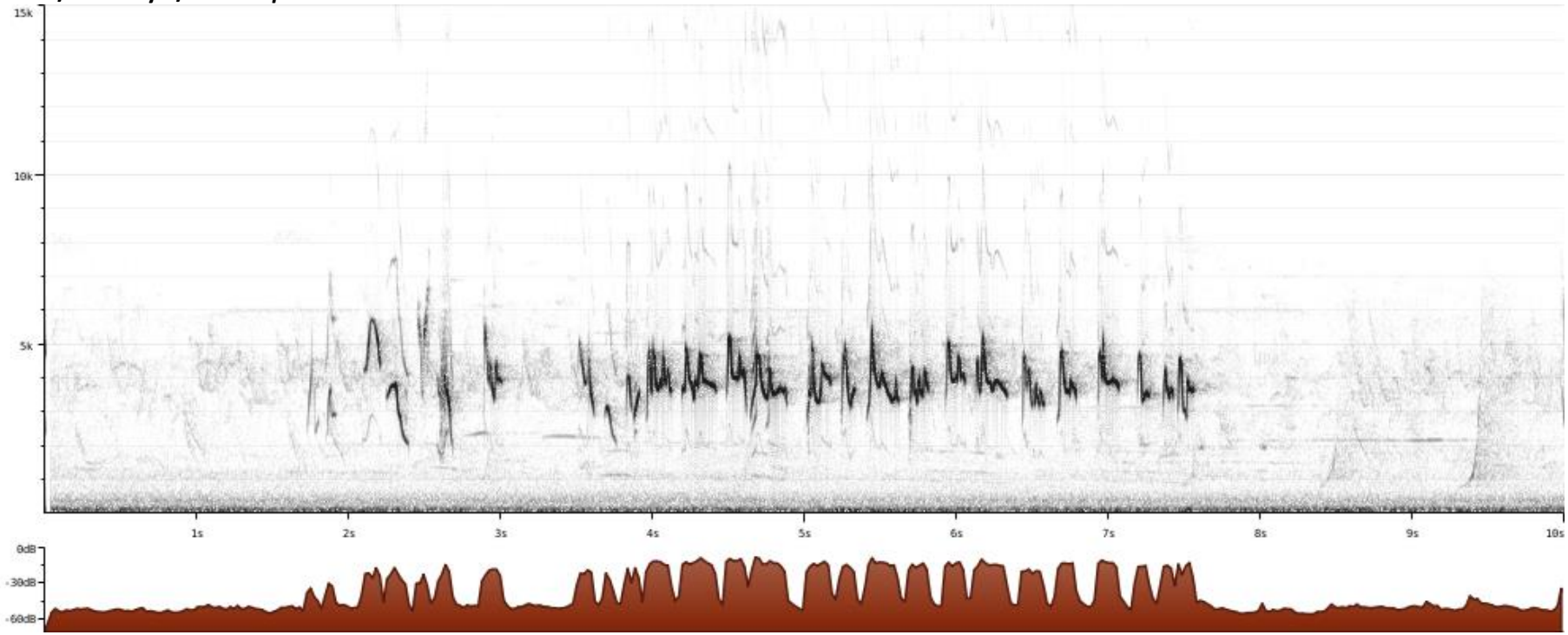
XC171909 / Australian Golden Whistler / Pachycephala pectoralis



http://www.xeno-canto.org/171909 / http://www.xeno-canto.org/sounds/uploaded/EHGWCIGILC/ffts/XC171909-large.png

Recordist	**Marc Anderson**
Date	2013-07-16
Time	10:00
Latitude	-28.4075
Longitude	153.0854
Location	**Border Ranges National Park (near Border Ranges), New South Wales**
Country	Australia
Elevation	800 m
Background	**White-browed Scrubwren** (*Sericornis frontalis*)

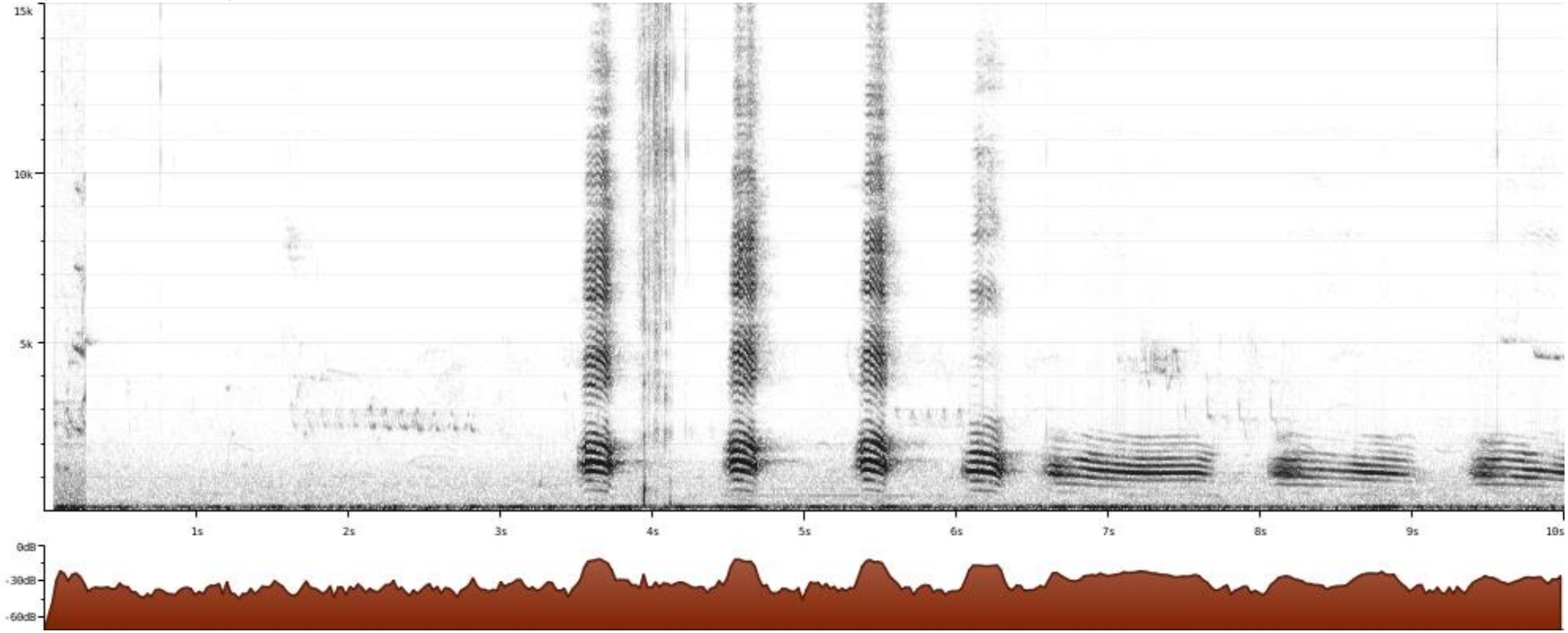
XC157483 / Silvereye / *Zosterops lateralis*



http://www.xeno-canto.org/157483 / http://www.xeno-canto.org/sounds/uploaded/UXGZWVYDFE/ffts/XC157483-large.png

Recordist	Fernand Deroussen
Date	2011-10-25
Time	05:00
Latitude	-26.4154
Longitude	153.0839
Location	Noosa Heads, Queensland
Country	Australia
Elevation	10 m
Background	non

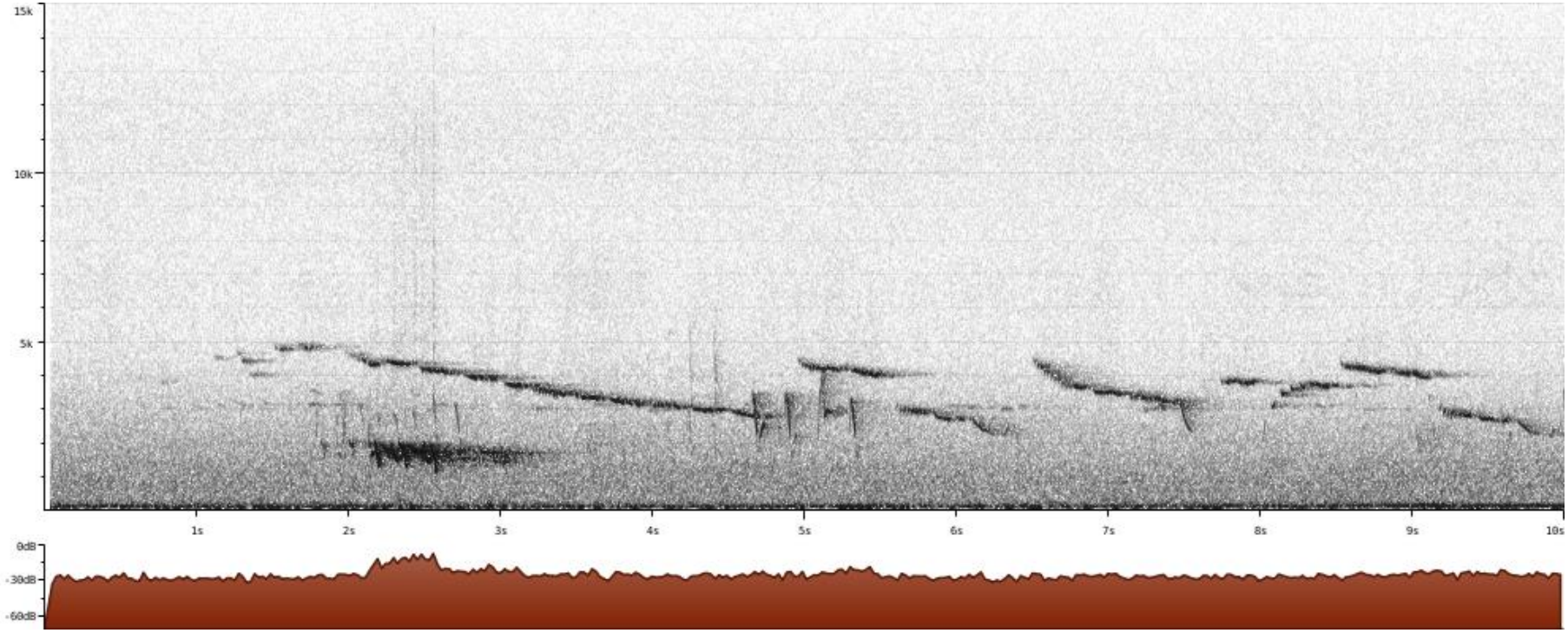
XC37368 / Torresian Crow / Corvus orru



http://www.xeno-canto.org/37368 / http://www.xeno-canto.org/sounds/uploaded/SILWLBBIFA/ffts/XC37368-large.png

Recordist	**Peter Woodall**
Date	2009-08-03
Time	10-07
Latitude	-27.4787
Longitude	153.1106
Location	**Minnippi Wetlands, Brisbane**
Country	Australia
Elevation	50 m
Background	none

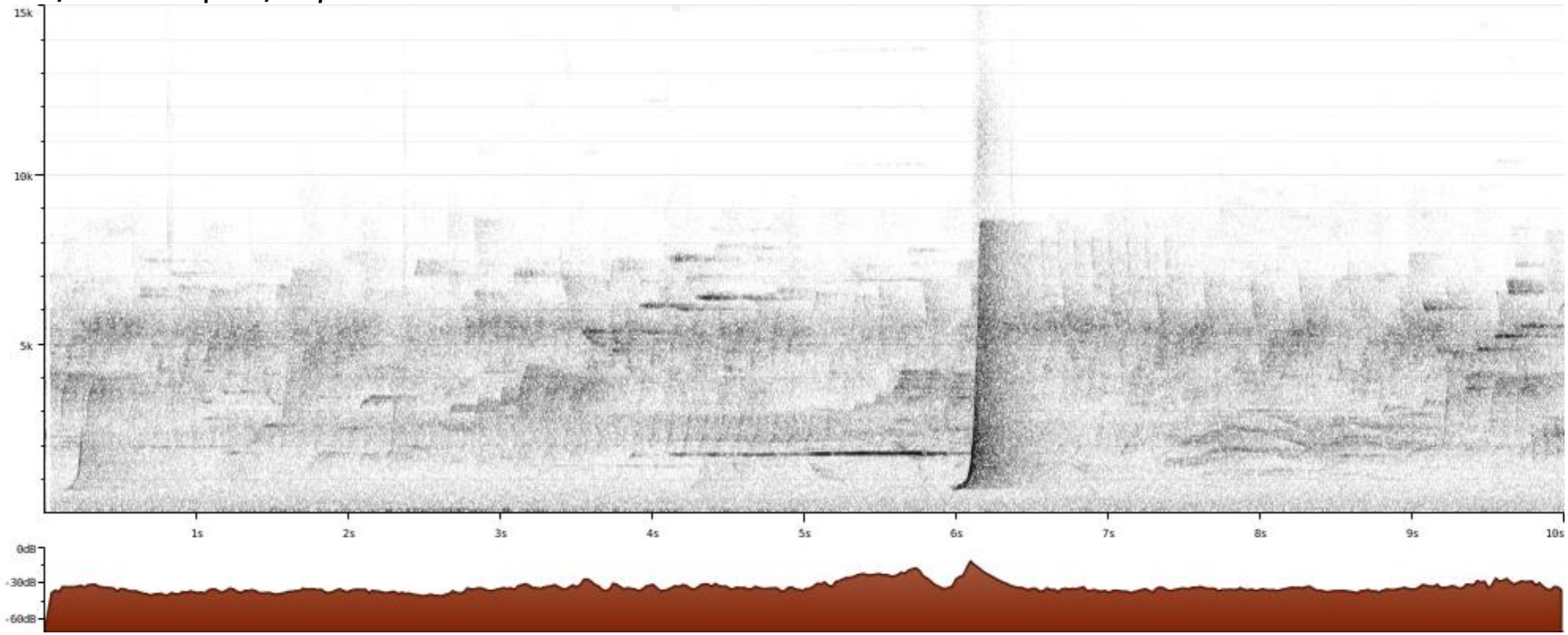
XC33176 / White-throated Gerygone / *Gerygone olivacea olivacea*



http://www.xeno-canto.org/33176 / http://www.xeno-canto.org/sounds/uploaded/XTVEPHMPPJ/ffts/XC33176-large.png

Recordist	**Niels Krabbe**
Date	2006-11-17
Time	09:42
Latitude	-27.3834
Longitude	152.9167
Location	**Brisbane Forest Park, c.20 km W Brisbane, Queensland**
Country	Australia
Elevation	130 m
Background	**Olive-backed Oriole** (*Oriolus sagittatus*)

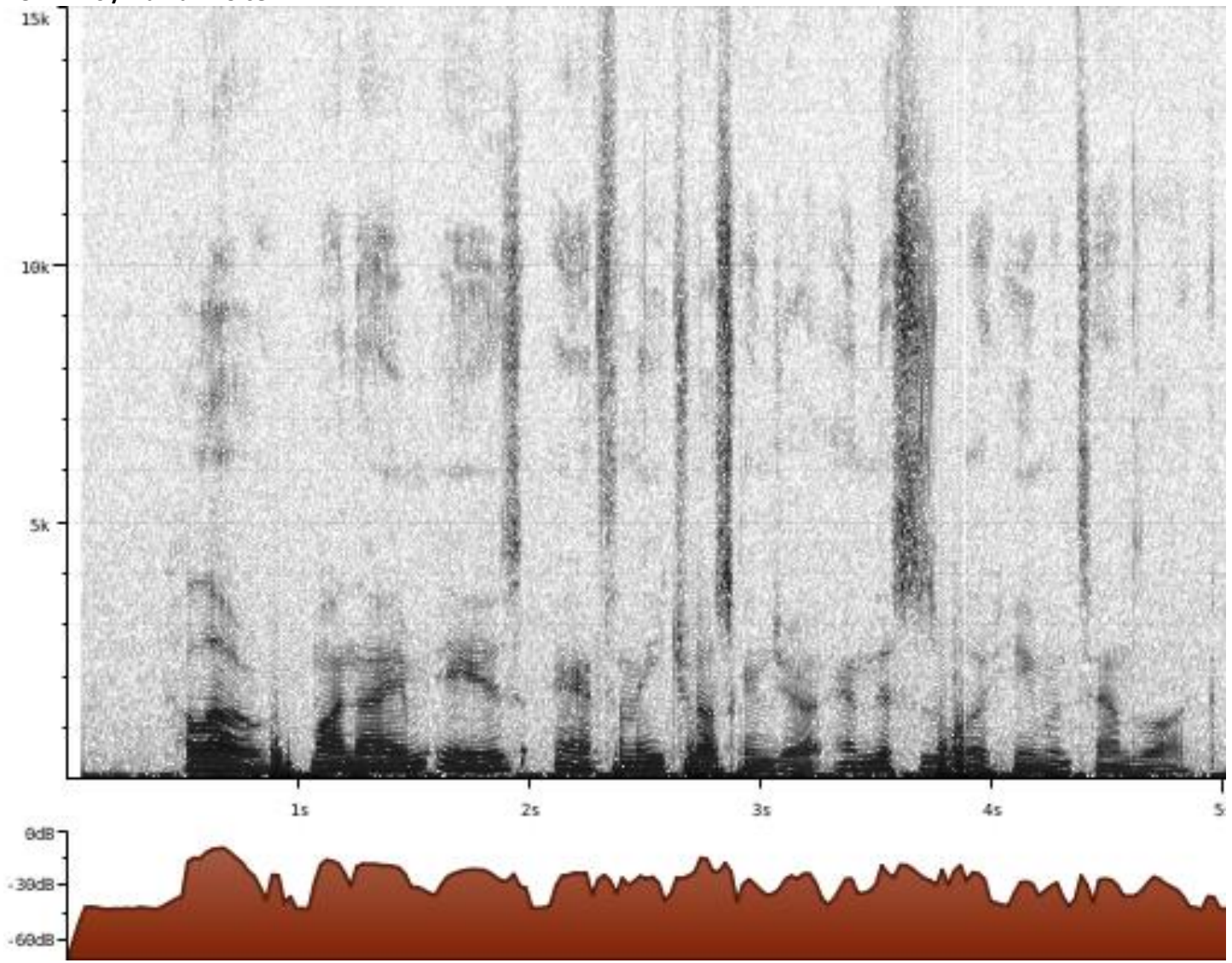
XC155100 / Eastern Whipbird / *Psophodes olivaceus*



http://www.xeno-canto.org/155100 / http://www.xeno-canto.org/sounds/uploaded/UXGZWVYDFE/ffts/XC155100-large.png

Recordist	**Fernand Deroussen**
Date	2011-10-22
Time	06:30
Latitude	-28.2287
Longitude	153.136
Location	**Lamington National Park (near O'reilly), Queensland**
Country	Australia
Elevation	900 m
Background	none

XC174426 / Human Voice



Recordist	Anthony Truskinger
Date	2014-04-14
Time	22:00
Latitude	-27.4774
Longitude	153.0274
Location	Brisbane, Queensland
Country	Australia
Elevation	20 m
Background	none

http://www.xeno-canto.org/174426 / http://www.xeno-canto.org/sounds/uploaded/OLTVFBWNZM/ffts/XC174426-large.png

Fin.